Advancing lentiviral gene therapy vectors for -thalassaemia

Stephanou, Coralea

Awarding institution:
King's College London

Advancing lentiviral gene-therapy vectors for β-thalassaemia

BY

CORALEA STEPHANOU

A thesis submitted for the degree of Doctor of Philosophy

2015

1st supervisor: Dr. Michael Antoniou
Division of Genetics and Molecular Medicine
Department of Medical and Molecular Genetics
Guy's Hospital – King's College London

2nd supervisor: Dr. Marina Kleanthous
Molecular Genetics Thalassaemia Department
The Cyprus Institute of Neurology and Genetics

Declaration

I, Coralea Stephanou, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Background: β-thalassaemia is a potentially lethal hereditary anaemia, caused by reduced or absent expression of HBB polypeptide chains of adult haemoglobin (HbA: $\alpha_2\beta_2$). Current curative treatment options are limited to few patients, while alternative, chronic palliative therapy consisting of frequent transfusions coupled with iron chelation therapy, is costly and reduces patients' quality of life. Increased levels of fetal haemoglobin (HbF: $\alpha_2\gamma_2$) were shown to lessen the severity of β-thalassaemia, highlighting the therapeutic potential of a gene-therapy-mediated increase in *HBG1* and *HBG2* (*HBG*) expression.

Aims: The aim of this project was to develop a new generation of lentiviral vectors (LVs) in order to deliver a combinatorial gene therapy for β-thalassaemia (and sickle cell disease). LVs were designed to enhance HbF levels by engineering the HBB-expressing GLOBE LV to co-express short-hairpin RNAs (shRNAs) that will post-transcriptionally knockdown repressors of *HBG* expression, such as BCL11A. Conceptually, it was envisaged that the combination of LV-derived HBB plus increased endogenous HBG chains would achieve levels of total haemoglobin that would reproducibly be in the range (~50% of normal values) to be curative for β-thalassaemia major (and sickle cell disease).

Methodology/Results: A series of modified LVs with simple shRNAs or microRNA-adapted shRNAs (shRNA-miRs) inserted within the second intron of *HBB* (βIVS-II) within GLOBE, were produced and functionally tested in erythroid tissue culture cell lines (K562, HEL, MEL) and in a human primary erythroid cell culture model system. Early experiments were conducted using shRNA-bearing vectors targeting mRNA of the GFP reporter to rapidly establish proof-of-principle for the vector design. Technical difficulties with the execution of experiments in GFP-expressing transgenic cell lines prompted the parallel use of BCL11A-targeting vectors aiming to induce *HBG* expression. The cell line work did not produce concrete findings, with every indication that these systems were unsuitable platforms for the intended analyses.

Lastly, vectors were functionally tested in primary erythroid cells derived from $CD34^+$ haematopoietic stem cells from both normal individuals and patients with β-thalassaemia. To

this end, extraction protocols of $CD34^+$ cells from peripheral blood samples were optimised, based on density-gradient centrifugation and simple ACK lysis of red blood cells. The effect of these protocols on cell viability, proliferation and differentiation was compared in standardised small-scale assays on semi-solid methylcellulose media. Standardising liquid cultures for large-scale $CD34^+$ cell expansion and differentiation, however, posed unexpected challenges and prompted the evaluation of three separate protocols before suitable cell numbers and differentiation levels were achieved. Using the established systems, the RNAi competency of stem-loop sequences in shRNAs was functionally validated by expression from the RNA-polymerase-III U6 promoter of control vectors. Vectors carrying the same basic shRNA cassettes in the GLOBE-encoded *HBB* βIVS-II showed no discernible effect on target genes. In contrast, subsequent preliminary analyses of second-generation shRNA-miR vectors demonstrated their RNAi activity, making them the subject of further investigation for combinatorial gene therapy for the haemoglobinopathies.

Acknowledgements

First, I would like to thank my academic supervisors Dr. Michael Antoniou and Dr. Marina Kleanthous for giving me the opportunity to work on such an interesting project in the field of haemoglobin gene therapy. Thank you for your guidance and encouragement over the last 4 years! Second, I would like to thank my project advisor, Dr. Carsten Lederer. I consider myself very fortunate to have worked with you. You have been a great help over these years, always supportive and available when I needed you. I appreciate your enthusiasm, diligence, and contribution of time and ideas to make this project, and at the same time my PhD experience, productive, stimulating, and enjoyable. Further to this, I would like to thank the Cyprus Institute of Neurology and Genetics (CING) for covering the research costs of this project, as well as my supervisors for their efforts to secure funding from the Research Promotion Foundation – Cyprus for the last two years of my PhD work.

I would like to thank my colleagues at the Molecular Genetics Thalassaemia Department (CING) for their help and support during my PhD period. I am thankful to Petros and Constantinos for their help throughout these years. It has been a great time being with you in lab 259α, and I am glad I had you to share my bench, worries and thoughts. Thank you also to Pavlos for his help with Western blotting and to Myria for her technical support in setting up the repository of leukocyte samples as a source of haematopoietic progenitors. From our external collaborators, I would like to thank Dr. Mario Amendola for donating the PGK-GFP transfer vector plasmid and auxiliary plasmids for LV production. I am also deeply grateful to Dr. Stefano Rivella and Dr. Laura Breda for providing training on 'Protocol C' for *in vitro* $CD34^+$ cell expansion and erythroid differentiation.

I would like to thank my friends and beloved ones for their unselfish help and unconditional love. Last but not least, a big thanks to my parents, Stephanos and Margarita, for always being my best supporter, for their love, patience and encouragement to pursue my interests…all of them! Thank you for everything xx

Abbreviations

The table does not list names of lentiviral open reading frames and official gene symbols and names, other than for genes included in **section 2.0**.

3'enh	β-globin 3' enhancer element
5-azacytidine	5-Aza-C
αMEM	α-Minimum Essential Medium
βp	β-globin promoter
M	molar
m	milli (10^{-3})
μ	micro (10^{-6})
n	nano (10^{-9})
A	adenine
AAV	adeno-associated virus
ACH	active chromatin hub
ACK	ammonium-chloride-potassium
AGM	aorta-gonad-mesonephros
AGO	argonaute protein
ampR	ampicillin resistance
APS	ammonium persulfate
BPB	bromophenol blue
BCL11A	B-cell CLL/Lymphoma 11A
B-EB	basophilic erythroblasts
BFU-E	burst forming unit-erythroid
BM	bone marrow
BMT	bone marrow transplantation
bp	base pair
BSA	bovine serum albumin
C	cytosine
CA	Cooley's Anaemia
CB	cord blood
cDNA	complementary DNA
CFC	colony-forming cell
CFU	colony-forming unit
CFU-E	colony-forming unit-erythroid

CFU-M	colony-forming unit-macrophage
CFU-G	colony-forming unit-granulocyte
CFU-GM	colony-forming unit-granulocyte, macrophage
CFU-GEMM	colony-forming unit-granulocyte, erythroid, macrophage, megakaryocyte
CFU-S	colony-forming unit spleen
cHS4	chicken hypersensitive site-4
CLP	common lymphoid progenitors
CMP	common myeloid progenitors
CMV	cytomegalovirus
cPPT	central polypurine tract
CRISPR	clustered regularly interspaced short palindromic repeats
C_T	threshold cycle
DC	dendritic cell
dH_2O	distilled water
DMEM	Dulbecco's modified eagle medium
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dpc	days post-conception
DSB	double-strand breaks
dsDNA	double stranded DNA
dsRNA	double-stranded RNA
dsRBD	dsRNA-binding domain
E	PCR efficiency
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EMH	extramedullary haematopoiesis
Em_{max}	emission maximum
EP	erythrocyte progenitor
EPO	erythropoietin
EpoR	erythropoietin receptor
ES	embryonic stem
Ex_{max}	excitation maximum
FACS	fluorescence-activated cell sorter
FBS	foetal bovine serum
FDA	U.S. Food and Drug Administration
Flt3L	Flt3 ligand
FSC	forward scatter
FW	forward
G	guanine

GAPDH	glyceraldehyde 3-phosphate dehydrogenase
G-CSF	granulocyte colony-stimulating factor
gDNA	genomic DNA
GFP	green fluorescent protein
GMP	granulocyte-monocyte progenitor
GP	granulocyte progenitor
GpA	glycophorin A
GT	gene therapy
GvHD	chronic graft-versus-host disease
GWAS	genome-wide association studies
Hb	haemoglobin
HbA	adult haemoglobin
HBA	human α-globin genes
HBB	human β-globin gene
Hbbh$_0$	murine βh$_0$-globin gene
Hbbh$_1$	murine βh$_1$-globin gene
Hbb1	murine β^{maj}-globin gene
Hbb2	murine β^{min}-globin gene
HBD	human δ-globin gene
HbE	embryonic haemoglobin
HBE	Human ε-globin gene
Hbe	murine εy-globin gene
HbF	foetal haemoglobin
HBG	human γ-globin genes
HBG1	human ^{A}γ-globin gene
HBG2	human ^{G}γ-globin gene
HBS	hepes buffered saline
HBZ	Human ζ-globin gene
Hbb$^{th3/+}$	murine thalassaemic
HDR	homology-directed repair
HEK 293T	human embryonic kidney 293T
HEL	human erythroleukaemia
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HIV	Human immunodeficiency virus
HLA	human leukocyte antigen
HMBA	hexamethylene bisacetamide
HMG	high mobility group
HPFH	hereditary persistence of fetal haemoglobin
hHPFH	heterocellular HPFH
pHPFH	pancellular HPFH

HS	hypersensitive site
HSC	haematopoietic stem cell
HSCT	haematopoietic stem and progenitor cell transplantation
HSPC	haematopoietic stem and progenitor cell
HU	hydroxyurea
IL3	Interleukin 3
IMDM	Iscove's Modified Dulbecco's Medium
IN	integrase
iPS	induced pluripotent stem
IVS	intervening region
kanR	kanamycin resistance
kb	kilobases
kDa	kilo Dalton
KLF1	Kruppel-like factor 1
KR	c-kit receptor
LB	lysogeny broth
LCR	locus control region
LF	large fragment
Lin	lineage markers
LM-PCR	ligation-mediated polymerase chain reaction
LT-HSCs	long-term-repopulating haematopoietic stem cells
LTR	long terminal repeat
LV	lentiviral vector
MacP	macrophage progenitor
MACS	magnetic-activated cell sorting
MARE	MAF-responsive element
MCM	methylcellulose media
MEL	murine erythroleukemia
MEP	megakaryocyte-erythrocyte progenitor
MFI	mean fluorescence intensity
miRNA	microRNA
MkP	megakaryocyte progenitor
MLV	Moloney murine leukemia virus
MMP	multipotent progenitor
MNC	mononuclear cell
MOI	multiplicity of infection
MPP	multipotent progenitors
MSC	mesenchymal stem cell
NC	nucleated cell
NES	nuclear export sequence

NHEJ	non-homologous end joining
NK	natural killer cell
NOD-SCID	nonobese diabetic/severe combined immunodeficiency
NSG	NOD-scid IL2rγ^{null}
NTC	non-template control
NuRD	nucleosome remodelling and histone deacetylase
OC-EB	orthochromatophilic erythroblasts
OD	optical density
ORF	open reading frame
PB	peripheral blood
mPB	cytokine-mobilized peripheral blood
PBMC	peripheral blood mononuclear cell
PBSC	peripheral blood stem cell
PBS	phosphate buffered saline
PC-EB	polychromatophilic erythroblasts
PCR	polymerase chain reaction
PEB	proerythroblast
PEG	polyethylene glycol
PGK	phosphoglycerate kinase 1
piRNA	PIWI-interacting RNA
PNC	polymorphonuclear cell
PR	protease
P/S	penicillin/streptomycin
PVDF	polyvinylidene difluoride
qPCR	quantitative polymerase chain reaction
R^2	square of Pearson's correlation coefficient
RBC	red blood cell
RIPA	radioimmunoprecipitation assay buffer
RISC	RNA-induced silencing complex
RITSC	RNA-induced transcriptional silencing complex
RNA	ribonucleic acid
RNAi	RNA interference
rRNA	ribosomal ribonucleic acid
RNA-pol	RNA-polymerase
RPMI	Roswell Park Memorial Institute medium
RRE	REV response element
RSV	Rous sarcoma virus
RT	reverse transcriptase
RTemp	room-temperature
RT-qPCR	reverse-transcriptase quantitative polymerase chain reaction

RP-HPLC	reversed-phase high performance liquid chromatography
RV	reverse
RV	γ-retroviral vector
SA	splice acceptor site
SCD	sickle cell anaemia
SCF	stem cell factor
SCID	severe combined immunodeficiency
SD	splice donor site // standard deviation (context-dependent)
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SF	short fragment
shRNA	short-hairpin RNA
shRNA-miR	microRNA-adapted shRNA
SIN	self-inactivating
siRNA	small interfering RNA
SNPs	single nucleotide polymorphisms
snRNPs	small nuclear ribonucleoprotein
SOX6	SRY (sex determining region Y)-box 6
SSC	side scatter
ST-HSCs	short-term haematopoietic stem cells
SV40	Simian vacuolating virus 40
T	thymine
T87Q	codon 87 change from threonine to glutamine
TALENs	transcription activator-like effector nucleases
TBE	tris-borate-EDTA
TEMED	tetramethylethylenediamine
TFA	trifluoroacetic acid
TGS	transcriptional gene silencing
THPO	thrombopoietin
TI	thalassaemia intermedia
TM	thalassaemia major
T_m	melting temperature
Transgenic HBB	vector-derived human β-globin
TU	transduction units
UTR	untranslated region
UV	ultra-violet
V	Volts
VCN	vector copy number
VSV-G	G glycoprotein of vesicular stomatitis virus
WBC	white blood cells

WHO	World Health Organization
WPRE	woodchuck hepatitis virus posttranscriptional regulatory element
XC	xylene cyanol
ZFs	zinc-fingers
Ψ	lentiviral packaging signal // pseudogene (context-dependent)
Δ	Deletion
v/v	volume per volume

Table of Contents

1.0 Introduction

1.1 Blood and haemoglobin

Blood consists of erythrocytes, leukocytes, platelets and nutrients that constantly circulate the body. It mainly functions as the body's transport system, supplying tissues with nutrition and oxygen and collecting metabolic waste and carbon dioxide for excretion. Additionally, blood has a major role as the body's defence system against infection. Therefore, regular blood testing serves an important tool one can use to detect medical abnormalities in order to prevent or manage a disease. The diagnosis of haemoglobinopathies follows a three-stage approach involving a complete cell count test to confirm anaemia, special haematological tests to identify disease severity [1], and deoxyribonucleic acid (DNA) mutation analysis to infer genetic causation. One of the main diagnostic parameters for anaemia-related diseases is the haemoglobin content of the erythrocytes.

Haemoglobin (Hb) is the oxygen transporter of erythrocytes. It was first described in 1840 by Friedrich Ludwig Hünefeld in samples of earthworm blood, held under two glass slides, as small plate-like crystals. In 1864, Felix Hoppe-Seyler named these crystals as 'haemoglobin' for the 'colorant substance of blood' [2]. The three-dimensional structure of haemoglobin was solved using X-ray chrystallography by Max Perutz in 1959, a discovery which earned him the Nobel Prize (1962) in chemistry. The structure is almost spherical with a diameter of 5.5 nm [3, 4]. It is made up of four polypeptide chains, two α-like chains and two β-like chains, each with eight helical segments, packed together in a tetrahedral array. Each chain consists of a heme group with a single oxygen binding site. The heme molecule is composed of a porphyrin ring and an iron atom ligand (Fe^{2+}) bound in the centre (**Figure 1.1**). When oxygen binds the heme, the iron atom adopts a ferrous (Fe^{3+}) oxidative state at which it can no longer bind oxygen. The binding of one oxygen molecule will trigger complex conformational changes within the protein in order to shield neighbouring heme subunits from becoming 'accidentally' oxidized while at the same time increasing their affinity for binding oxygen (relaxed Hb state). This cooperative behaviour between subunits ensures that all oxygen-binding sites are occupied. In a similar fashion, losing one oxygen molecule will trigger structural changes which lower the affinity for the remaining oxygen molecules (tense Hb state). Hb is a well-defined allosteric protein of which

transformation between a relaxed and tense structural state is guided by the oxygen pressure in the surrounding environment. The cooperative binding of oxygen by haemoglobin ensures that 1.7 times more oxygen is delivered from the lungs to the tissues as it would if the oxygen-binding sites were independent [5].

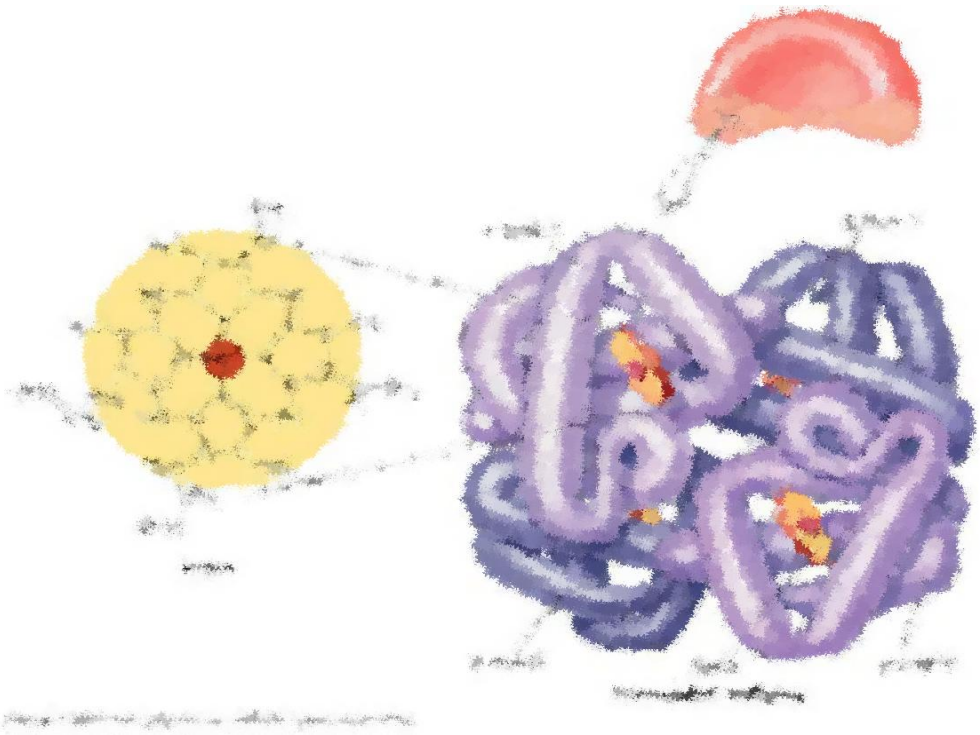


Figure 1.1. Diagrammatic illustration of the structure of haemoglobin (HbA: $\alpha_2\beta_2$). Reproduced from http://www.namrata.co/structure-of-haemoglobin-an-overview/

1.2 Erythropoiesis

Erythropoiesis is a multistep developmental process which involves differentiation of multipotent haematopoietic progenitors into erythroid-committed blasts and terminates with the production of mature erythrocytes (red blood cells; RBCs). During progressive stages of erythroid cell maturation, blast cells exhibit gradual loss of their multipotency while increasing lineage restriction, undergo extensive cellular and nuclear remodelling thus resuming distinct morphological characteristics, and ultimately extrude their nucleus to form reticulocytes which then become mature erythrocytes in the circulation [6]. Regulators of erythropoiesis include transcription factors (e.g., GATA1, SCL/TAL1, LMO2, LDB1, KLF1), miRNAs, and histone

modifications (see [7, 8]). Mammalian erythropoiesis exists in two developmentally and morphologically distinct forms of erythroid cell production, the primitive (embryonic) and the definitive (adult) stages (**Figure 1.2**).

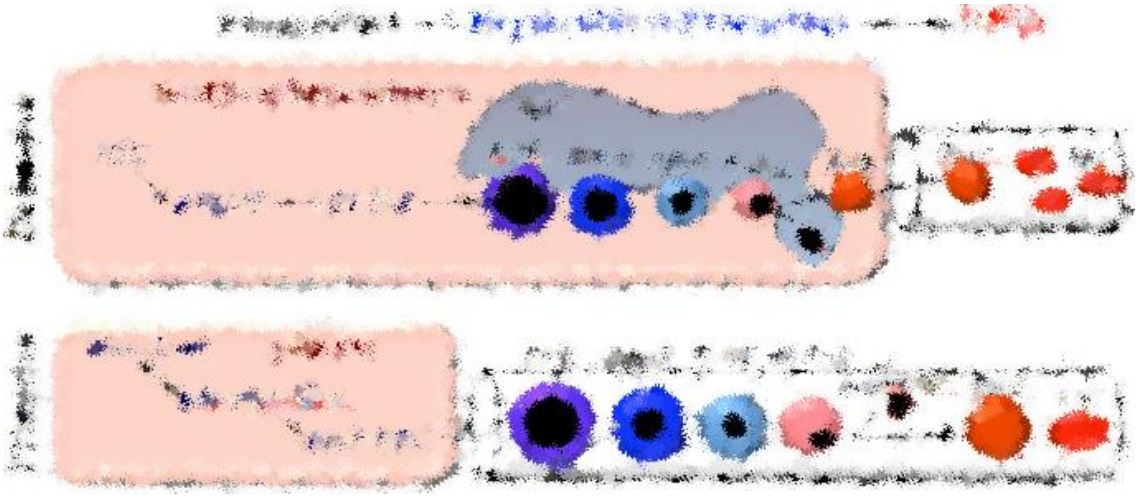


Figure 1.2. Schematic illustration of primitive and definitive erythropoiesis. Embryonic (primitive) and adult (definitive) erythropoiesis involves maturation of cells through three compartments; the multilineage progenitors, the unilineage precursors and the mature functional-end red blood cells (RBCs). Primitive erythropoiesis occurs just once during embryogenesis from mesoderm cells whereas definitive erythropoiesis occurs continuously throughout adult life from haematopoietic stem cells (HSCs). Reproduced from [9].

1.2.1 Haematopoiesis – general overview

Erythropoiesis stems from a dynamic, orderly developmental process called haematopoiesis which operationally describes the commitment and differentiation of haematopoietic stem cells (HSCs) to produce myeloid and lymphoid lineage-restricted cells, collectively forming all blood cell types.

Stem cells were first functionally characterized in 1960s by Till and McCulloch [10] who demonstrated the ability of syngeneic mouse bone marrow (BM) grafts to form cellular colonies in the spleens of lethally irradiated recipient mice. Analysis of these colonies, called colony-forming units-spleen (CFU-S), identified the presence of a multilineage cell progenitor of which a subset could reform CFU-S in secondary transplanted mice. This was a pioneering study

showing the ability of stem cells for self-renewal (generation of daughter cells without differentiation) and multipotentiality (differentiation into all types of functional blood cells) [11, 12]. Subsequent studies identified stem cells as the source of virtually all highly-differentiated cells and classified them into three categories based on their potency; i) pluripotent cells (e.g., embryonic stem cells) with ability to differentiate into any type of mature cell, including the earliest of tissues, the trophectoderm, ii) multipotent cells (e.g., HSCs) with ability to form multiple differentiated cells belonging to the myeloid and lymphoid lineages, iii) unipotent cells (e.g., germline (sperm/egg) stem cells) with ability to produce a single cell type [13].

Haematopoiesis has a hierarchical structure in which the HSCs represent the founding cell population (**Figure 1.3**), which gives rise to all blood lineages of the adult organism via a series of well-defined proliferation and commitment steps during the course of haematopoietic development. Haematopoiesis is sustained throughout adult life by maintaining a balance between three cellular fates; self-renewal, commitment to differentiate, and death [13]. Adult HSCs reside within their microenvironmental niche in a quiescent state and divide infrequently solely to self-renew and expand without compromising their long-term proliferation potential. They possess multipotency and display long-term repopulating potentials (LT-HSCs), which are essential to sustain long-term engraftment and haematopoietic reconstitution following injury or BM transplantation. They have a relative frequency of 1:10000 in the BM, from which a subset divides to produce multipotent progenitor cells (MMP), which retain full lineage potential but a limited capacity for self-renewal. These are called short-term repopulating progenitors (ST-HSCs) and have a relative frequency of 1:2000 in the BM. Their transient contribution to haematopoiesis was documented in transplantation studies, where ST-HSCs were associated with early phase engraftment and accelerated haematopoietic reconstitution. MMPs give rise to the oligopotent progenitors which possess more restricted developmental potential, which in turn give rise to more lineage-restricted progenitors from which all terminally-differentiated blood cells arise. At every step of the haematopoietic hierarchy, cells may trigger their death as a means to restore balance by regulating cell numbers and eliminating mutant clones generated during frequent proliferation [8, 11, 13-16].

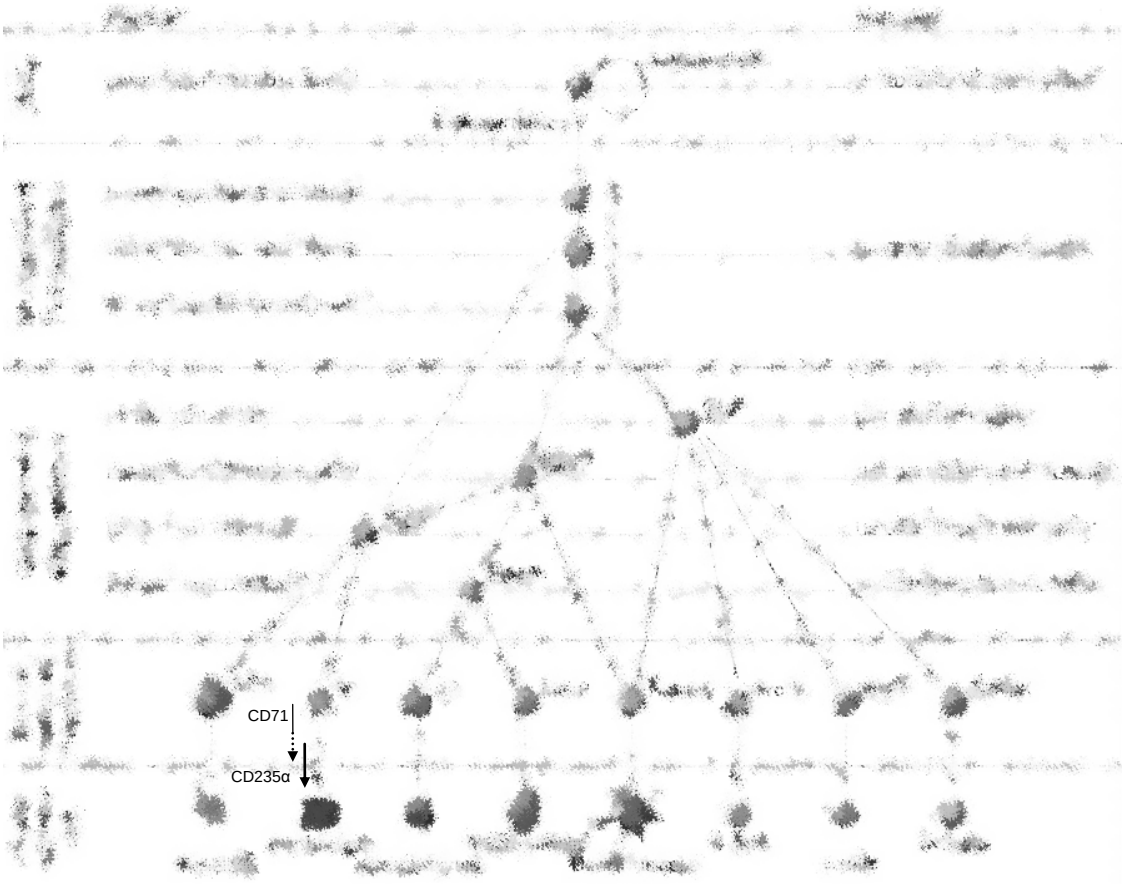


Figure 1.3. Hierarchical organisation of haematopoiesis. HSCs head up the hierarchical model of haematopoiesis, and possess both self-renewal and multilineage potential. Throughout differentiation, successive progenitors lose proliferative and lineage potential, until they become mature functional cells of a certain lineage. The cell surface phenotype of these cell types in the mouse and human system is shown. The temporal expression of erythroid-specific CD71 and CD235α surface markers on erythroblasts at different maturation stages is detailed in **Figure 1.4**. HSC: haematopoietic stem cell, CLP: common lymphoid progenitor, CMP: common myeloid progenitor, GMP: granulocyte/macrophage progenitor, MEP: Megakaryocyte/erythrocyte progenitor, MkP: Megakaryocyte progenitor, EP: erythrocyte progenitor, GP: granulocyte progenitor, MacP: macrophage progenitor, DC: dendritic cell, NK: natural killer, Lin: lineage markers. Reproduced from [12].

1.2.2 HSC microenvironmental niche

HSCs depend on their microenvironment, the niche, for regulation of survival, expansion and differentiation. During ontogeny, haematopoiesis is established in different anatomical sites with every developmental niche comprised of a specific signalling pathway or adhesion molecule which allows HSC to maintain contact with the niche cells and with the extracellular matrix, which in turn influences their lineage fate [11, 13]. The sequential sites of haematopoiesis include the embryonic yolk sac and surrounding aorta-gonad-mesonephros (AGM) region, the fetal liver, and finally the adult BM. Embryonic haematopoiesis is transient and rapidly replaced by adult-like haematopoiesis. The fetal liver does not support *de novo* HSC production; instead it serves as a site for HSC expansion and proliferation. While fetal HSCs are in cycle, in the BM, HSCs are mainly quiescent and produced throughout adult life [17].

1.2.2.1 Bone marrow HSC niche

The BM represents the most complex of the haematopoietic sites, comprised of two niches, the endosteal and the vascular. The endosteal niche consists of the endosteal surface of the bone and cells of osteoblastic lineage. A particular subpopulation of osteoblastic cells, called osteoblasts, has been implicated in the regulation of HSC fate through an array of secreted and cell-to-cell adhesion molecules. The osteoblasts arise from mesenchymal stem cells (MSCs) and can be identified by the production of CXCL12 (also termed SDF1; stroma-derived factor 1) and CXCR4 factors which regulate HSC trafficking and localisation in the marrow niche. Other cytokines and adhesion molecules (e.g., SCF, THPO, MPL, ANGPT1, and TIE2 kinase) have been shown to regulate HSC-osteoblast interaction and, specifically, HSC maintenance and *in vivo* quiescence. Specialised endosteal macrophages, called osteoclasts, have been shown to egress HSCs from the BM in addition to their inherent properties of bone resorption [18]. Furthermore, the sinusoidal vasculature, which is in close proximity to the endosteal surface, has been suggested as a niche for HSC maintenance and regeneration since the majority of HSCs were shown to reside within the sinusoids (vascular spaces lined by a fenestrated endothelium) [19]. It was demonstrated that the BM endothelial cells secrete angiocrine factors involved in the reconstitution, expansion, and maintenance of HSCs. Therefore the BM

vasculature is pivotal for HSC recovery from myeloablative injuries and following BM transplantation. The HSCs are specifically located adjacent to reticular cells expressing high levels of CXCL12 (hence the name CAR cells), which surround the sinusoid endothelium. CAR cells were found to be involved in HSC proliferation (maintained in undifferentiated state) and may constitute adipo-osteogenic precursors, which contribute to haematatopoietic regulation in both the endosteal and the vascular niches [20]. In addition, the protein and chemical composition of the osteoblast extracellular matrix has also been implicated to influence HSC fate. For example, the SPP1 protein (or osteopontin) was shown to influence HSC numbers and function [18]. Moreover, a number of adhesion molecules have been shown to connect HSCs to their haematopoietic microenvironment such as via integrin α4β1 expressed by the extracellular matrix or VCAM1 expressed on the surface of endothelial and/or osteoblastic cells, and also to interlink with signalling cascades (e.g., the Wnt, Notch, and hedgehog signalling pathways), thus, further adding to the complexity of molecular interactions underlying HSC regulation in the BM niche [18]. A unique characteristic of the BM niche is the capacity to influence HSC mobilisation (leaving the BM) and homing (returning to the BM), mainly through the SDF1-CXCR4 signalling axis. This feature has been successfully exploited therapeutically during BM transplantation [13].

1.2.2.2 The erythroblastic island

Erythroblastic islands are microniches which support RBC production and are found throughout the BM. They mainly consist of macrophages, which physically and chemically provide the microenvironment for erythroblast development. These so-called island macrophages are subsets of central resident macrophage cells (CD11c+CD18+) residing in the haematopoietic tissue [21]. Their main functions include iron recycling from senescent RBCs, phagocytosis of extruded nuclei during RBC enucleation, and supporting proliferation and survival of maturing erythroblasts. The erythroblast-macrophage contact is supported by adhesion molecules and paracrine factors secreted within the erythroid niche. Examples include the adhesion molecules VCAM1 and ICAM4/αV, and integrin α4β1 for maintaining erythroblastic island integrity, as well as transcription factors (e.g., GATA1) and extracellular matrix proteins (e.g., fibronectin, laminin)

for regulating erythroid differentiation. An important factor is the molecular mediator Emp (erythroblast macrophage protein) in the absence of which ensuing severe anaemia can result in perinatal death. These and additional factors are reviewed in [22]. Island macrophages have been implicated in erythroid proliferation and differentiation during stress erythropoiesis. The therapeutic benefit associated with targeting macrophages in order to normalize erythropoiesis in the context of polycythemia vera and β-thalassaemia (i.e., decrease erythropoietic activity in disorders characterized by enhanced erythropoiesis) has been reported [23, 24].

1.2.2.3 Extramedullary haematopoiesis

Extramedullary haematopoiesis (EMH) is the production of RBCs outside the medullary spaces of BM, commonly affecting the reticuloendothelial system which involves the liver, spleen and lymph nodes. EMH occurs when RBC production in the BM is insufficient to meet demand, and is usually associated with acquired BM replacement disorders (e.g., myelofibrosis, leukemia, lymphoma) or congenital haemoglobinopathies (e.g., thalassaemia, sickle cell anaemia (SCD), spherocytosis). The BM vascular niche is important for establishing and sustaining haematopoiesis in the extramedullary tissues. Indeed, circulating haematopoietic precursors provide an immediately available pool that can be recruited rapidly for EMH. A hypoxic environment favours the formation of optimal conditions for preferential homing and expansion of HSCs at extramedullary sites. The development of EMH mainly depends on stromal proliferation mediated by resident macrophages and the activation of cytoadhesive and/or chemokine signalling pathways (e.g., the SDF1–CXCR4 signalling axis), which induce HSC mobilization and homing in hemic tissues [21].

1.2.3 Definitive erythropoiesis

The erythropoiesis differentiation pathway begins with the bipotent progenitor MEP, which differentiates in response to an erythropoietin (EPO) stimulus to produce the most immature erythroid-committed progenitor, the BFU-E (burst forming unit-erythroid), and which constitutes 0.04%–0.12% of BM stem cells. BFU-Es further differentiate to generate the latest immature definitive erythroid-committed progenitor, the CFU-E (colony forming unit-erythroid), which

constitutes 0.2%–0.6% of BM stem cells [25]. They express the CD34 marker, an O-glycosylated cell surface glycoprotein antigen common to all early progenitors, allowing for their isolation using anti-CD34 antibodies. CFU-Es have little in common with BFU-Es, despite having indistinguishable histologic characteristics. Instead, CFU-Es closely resemble features of mature erythroid precursor cells, including the absence of self-renewal, limited proliferative capacities, presence of erythroid-specific antigenic expression profiles (glycophorin A and transferrin receptor) and high sensitivity to EPO [26, 27]. Studies have shown an abundance of EPO receptors (EpoR) on CFU-Es such that CFU-E proliferation and consequently the output of mature erythrocytes, is modulated by the levels of EPO in the circulation. In cases of acute stress (e.g., anaemia) the reduction in oxygen levels triggers an increase of EPO concentration in the circulation, which in turn uncouples the self-renewal potential of CFU-Es and stimulates their terminal proliferation and differentiation. However, in cases of chronic erythroid stress, steady-state EPO-dependent CFU-E production is insufficient to correct the RBC deficiency therefore triggering the production of novel CFU-Es from BFU-Es through, as yet, unknown mechanisms. Notably, EPO is the key modulator of erythropoiesis but with the latter stages of erythroid maturation being EPO-independent as indicated by minimal-to-complete loss of EpoR expression with progression of erythroid differentiation [9, 25]. Together, BFU-Es and CFU-Es constitute the erythroid progenitor cell compartment and are defined by their ability to form colonies in cytokine-supplemented semi-solid methylcellulose media. BFU-Es are relatively quiescent (10%–20% in cycle) with a high proliferative capacity when triggered *in vitro*, producing erythroid bursts of >200 Hb-containing cells after 14 days in culture. However, mature CFU-Es have a higher mitotic rate with 60%–80% of cells in cycle and proliferate faster in culture to produce erythroid colonies within 7 days but which are smaller (8–64 cells/colony) compared to BFU-Es as they require fewer divisions to attain haemoglobinization [28]. Following on from the CFU-E erythroid stage, the succeeding blast cells assume morphologically identifiable changes and collectively form the erythroid precursor cell compartment, also called the ‘erythron’.

The earliest recognizable erythroid cell is the proerythroblast, which undergoes three mitoses to generate sequentially basophilic, polychromatophilic and orthochromatic erythroblasts (**Figure**

1.4). Each blast cell resumes distinct morphologic characteristics, which reflects its stage of maturation in a series of highly regulated developmental changes including gradual decrease in size, chromatin condensation and RNA content reduction, and accumulation of erythroid-specific proteins (e.g., haemoglobin) [6, 9, 29]. Their expression ratio follows a 1:2:4:8 doubling pattern with successive mitoses, which reflects the physiological progression to terminal erythroid differentiation [26]. The ordered differentiation through the four precursor stages occurs over four days [30], followed by expulsion of the condensed erythroblastic nucleus through the cell membrane, a complex process called enucleation, and production of reticulocytes and pyrenocytes. Erythroblast enucleation is supported by local macrophages residing erythroblastic islands. Reticulocytes have been proposed to secrete paracrine factors which disturb the reticulocyte–macrophage attachment, releasing the reticulocytes into the BM sinusoids and subsequently into the peripheral circulation [22]. Changes in adhesive interactions were also proposed to help disconnect the reticulocytes from the pyrenocytes ('extruded nuclei', a small nucleated cell with a rim of cytoplasm), a process which leads to the exposure of phosphatidylserine (apoptotic 'eat-me' flag) on the pyrenocyte cell membrane therefore causing their engulfment and phagocytosis by central macrophages [31, 32]. Reticulocytes undergo extensive processing, which results in the generation of mature circulating RBCs. This maturation process entails a 20% loss of plasma membrane surface area, cell volume reduction, cytoplasmic organelle loss through autophagy and exocytosis, and cytoskeletal remodelling, in order to generate the classic biconcave discoid shape of mature erythrocytes [33].

During terminal erythroid differentiation, 2 million non-nucleated reticulocytes are generated every second [27] in order to maintain a steady-state emanation of $2x10^{11}$ blood cells per day in the bloodstream [25]. Each erythrocyte has a mean lifespan of 120 days with a ±15% variation detected among normal individuals [7, 34] whereas, in the case of haemolytic anaemias, survival rates are considerably lower owing to cytoskeletal protein abnormalities which cause further membrane surface area loss. Haemolysis ensues with destruction of abnormal erythrocytes by splenic macrophages and may lead to severe anaemia if the balance between RBC destruction and production is not restored [35]. Naturally aging RBCs exhibit loss in

surface area and volume, which is consistent with a 15% loss of cellular haemoglobin through vesiculation [36]. A build-up of cytoskeletal compression will eventually restrict further loss of membrane area, causing accumulation of oxidative waste material (e.g., degraded haemoglobin) in the erythrocyte membrane, thus triggering band 3 clustering and IgG binding (regarded as senescence markers), as well as the exposure of the phosphatidylserine flag, ultimately leading to RBC removal [36, 37]. In addition, the oxidative damage will disturb cellular ionic homeostasis causing dehydration with debilitating effects on RBC deformability, blocking movement through the microcirculation and eventually triggering RBC destruction [38]. Given the relatively short life-span of RBCs, a high turnover rate requires the use of profound homeostatic control mechanisms, which are regulated within specialised microenvironmental niches, from embryo to adult, depending on the age of development.

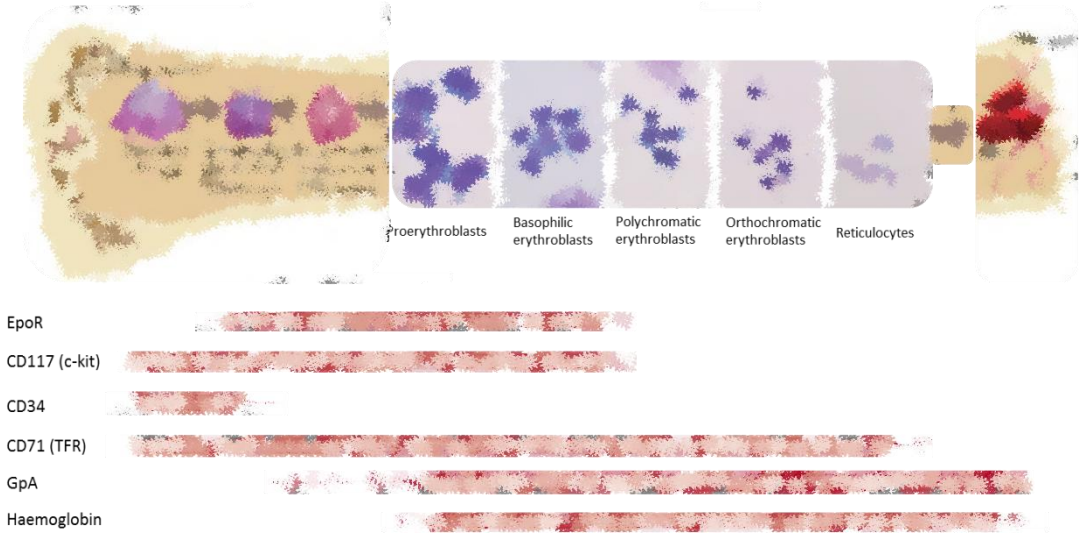


Figure 1.4. Stages of definitive erythroid differentiation. The diagram shows the various stages of human erythropoiesis, with representative images of erythroblast morphology on stained cytospins. Time dependent expression of various markers (EpoR, CD117, CD34, CD71, and GpA) and effector molecules (haemoglobin) is shown. Flow cytometry can be used to define the developmental stage in erythroid maturation. Modified from [25, 27, 39].

1.2.4 Primitive erythropoiesis

During embryogenesis, the primary site of haematopoiesis and origin of HSCs is the yolk sac. Briefly, *de novo* blood vessel formation (vasculogenesis) begins soon after gastrulation (E6.5 in mice) and subsequent to migration of mesodermal progenitor cells towards the extra-embryonic yolk sac where they proliferate to form angioblastic cords. The differentiation of angioblastic cords gives rise to 'blood islands' composed of primitive haematopoietic precursors and surrounded by a layer of endothelial precursors (angioblasts) [40]. Given that the haematopoietic and endothelial cell lineages exhibit an intimate spatial and temporal emergence, the presence of a common ancestral progenitor was suggested. Further studies in cultured ES cells identified the presence of a unique blast colony-forming cell (blast-CFC) with dual differentiation potential. The mesodermal cells give rise to a bipotent precursor called the haemangioblast with abilities to differentiate into both haematopoietic and endothelial cells [41, 42]. Primitive erythroblasts emerge (on E7.5 in mice and 30 days post-conception (dpc) in humans [43]) from unique progenitors termed primitive erythroid colony-forming cells (EP-CFC). Studies of the mouse conceptus revealed an increase in EP-CFC numbers by >7-fold at E8.0, after which they decreased to undetectable levels at E9.0. The narrow temporal wave of EP-CFC production reflects the transient nature of primitive erythropoiesis [41, 44].

In the human embryo proper, the para-aortic-splanchnopleura tissue, which is endowed with haematopoietic potential, develops into the dorsal aorta of the AGM region. The latter is widely recognized as the main intra-embryonic source of adult HSCs capable of long-term reconstitution [17, 20, 45], despite critical involvement of *RUNX1* expression in and possibly of cells from the developing extraembyonic mesoderm in the process [46, 47]. It is generally thought that definitive HSCs develop from endothelial-like precursors called haemogenic endothelial cells within the vascular endothelial aortic layers and mature in intra-aortic cell clusters budding from the luminal surface of the dorsal aorta (E27–E40 in human, E9.5–E11.5 in mice) [48, 49]. These HSCs express *RUNX1*, which favours haematopoietic development over other differentiation fates in the haemogenic endothelium [17, 40, 50]. Within the arterial vasculature, haematopoietic potential was also detected in vitelline and umbilical arteries, which

connect the aorta to the yolk sac and the placenta [13]. The placenta was considered as an additional extraembryonic site for *de novo* haematopoiesis (E8.5–E10.5 in mice), with chorionic and allantoic mesoderm constituents being regarded as the origin of HSCs in the tissue. Findings that these cells could provide haematopoietic repopulation of nonobese diabetic/severe combined immunodeficiency (NOD-SCID) mice further reinforced *in situ* origin of HSCs in the placenta [20]. Primitive erythroid precursors undergo progressive morphological changes to produce functional erythrocytes in a similar fashion as their definitive counterparts. Developmental changes include accumulation of haemoglobin, chromatin condensation and transcription inactivation, and reduction in size. Because primitive erythroblasts mature in the embryonic circulation, it is possible that they sustain cytoskeletal remodelling prior to enucleation and terminal maturation to RBCs [9]. Mature erythroid precursors were reported to enter the developing fetal liver (E11 in mice) prior to the migration of AGM-derived multilineage HSCs, which would form latter-stage definitive erythroid progenitor cells. Subsequently, haematopoietic precursors will seed the BM which forms the primary site of blood production in adults [45].

1.3 The β-thalassaemias

1.3.1 Disease and global distribution

The haemoglobin disorders are severe and common hereditary anaemias resulting from defects in globin expression, with about 7% of the world's population carrying a pathological haemoglobin gene [51]. Haemoglobinopathies in the majority of cases follow a recessive mode of inheritance with a 25% risk of an affected child being born with a potentially severe anaemia if couples are carriers of haemoglobinopathies. Many mutations have been characterized in the globin loci, a number of which cause defective globin production (the thalassaemia syndromes) and others which affect globin polypeptide chain structure and stability (the variant Hbs). The thalassaemias are classified according to the globin chains affected and the commonest types are the α- and β-thalassaemias. On a global scale, >700 variant Hbs have been identified from which the Hb C, Hb E, and Hb S are encountered at higher frequencies and pose a greater public health concern [51].

Haemoglobinopathies were originally described in the Mediterranean region and large parts of Asia and Africa, and many of the disease-causing mutations have spread throughout the world via international migration [52]. The global distribution of the thalassaemias and Hbs variants is shown in **Figure 1.5**. The carrier frequency for severe haemoglobinopathies throughout South-East Asia has been reported in the range of 5%–40%, reaching up to 70% regionally, and decreasing to 10%–20% in parts of Central Asia and the Indian sub-continent. Across the Arab nations, carrier frequencies were reported in the range of 5%–40% to as high as 60% in some areas of the population. Carrier rates were high (5%–30%) for parts of Africa, dropping to 5%–20% in the U.S.A. and Central America, and becoming less frequent (6%–10%) in Europe, which is mainly concentrated in Italy and Greece [52].

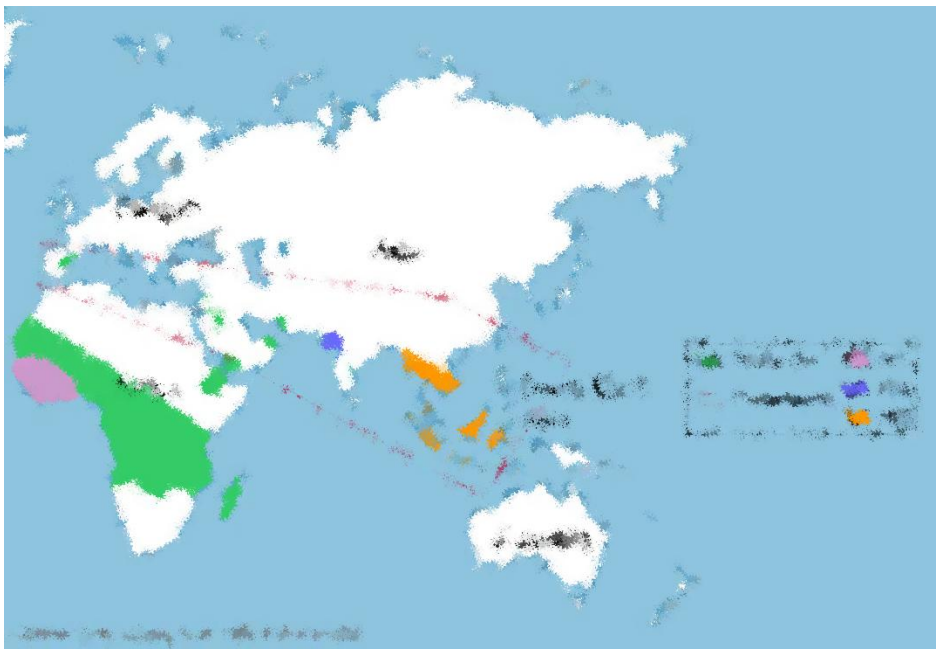


Figure 1.5. Global distribution of thalassaemia and Hbs variants (HbS, HbC, HbD, and Hb E). Reproduced from http://www.elu.sgul.ac.uk/rehash/guest/scorm/275/package/content/investigation.htm

In thalassaemias, the stoichiometry of the main adult haemoglobin molecule HbA ($\alpha_2\beta_2$) is affected by the absence or severely lowered expression of one of the constituting globin chains. Early cases of anaemia were reported in 1925 by Thomas Cooley and Pearl Lee in children of Italian origin and subsequently in parts of the Mediterranean region which gave the disease its name as 'thalassaemia' from the Greek word 'thalassa' which means 'sea' [53]. Generally, β-thalassaemia is in fact widespread, extending from the Mediterranean basin and parts of Africa throughout the Middle East, the Indian sub-continent, South-East Asia, the Malay Peninsula and into the Pacific Islands. The carrier frequency in these regions was estimated in the range of 1%–20%. The frequency of α-thalassaemia is higher (10%–20%) but shows a more restricted global distribution mainly in parts of the Mediterranean and South-East Asia and therefore is less of a global health problem than is β-thalassaemia [51, 53]. The high frequency of thalassaemias in regions endemic for *Plasmodium falciparum*, the parasite causing malaria, reflects a heterozygote survival advantage against severe malaria and a natural selection for preserving the disease-causing mutations in the population [54]. In addition, in communities where consanguineous marriage is traditional, thalassaemia prevalence is higher [55].

1.3.2 The genetics of β-thalassaemia

The severity of β-thalassaemia depends on the extent of imbalance in globin-chain production, with more than 200 disease-causing mutations described to date [55, 56]. The majority of mutations are single nucleotide substitutions within the β-globin gene (*HBB*) or its immediate flanking regions, small deletions and insertions of a few bases leading to a frameshift in mRNA coding capability. Large deletions are rare but can occur with an example being the Indian 619 bp deletion [56]. The mechanism by which these mutations affect *HBB* function can be at any level of transcription to RNA processing and translation of HBB mRNA. Other rare mutations include deletions of upstream regulatory elements (namely the βLCR), which results in down-regulation or complete absence of expression of the linked HBB-like family genes. Mutations that are physically linked to *HBB* form alleles of the *HBB* locus, and exhibit a wide spectrum of allelic heterogeneity between high-prevalence populations [55-57]. An updated description of existing mutations causing the thalassaemia syndromes, including their global distribution, can be found on the ITHANET portal at http://www.ithanet.eu/db/ithagenes.

Mutations that inactivate *HBB* and abolish HBB polypeptide production are called β^0-thalassaemia, whereas mutations that reduce HBB chain output are called β^+-thalassaemia. Heterozygosity for one β-thalassaemia allele leads to microcytic anaemia and elevated HbA_2 (3.5%–5.5%; increased interaction between the δ gene and upstream βLCR [57]), a phenotype characteristic of β-thalassaemia trait. Contrastingly, heterozygosity for a mild or 'silent' β-thalassaemia allele retains residual β-globin output from the affected β-locus, which does not result in any haematological abnormalities. Homozygosity for β-thalassaemia leads to severe transfusion-dependent anaemia characteristic of thalassaemia major (TM). In addition, compound heterozygotes for one β-thalassaemia allele and the Hb E variant (a high prevalence genotype in South-East Asia and the Indian sub-continent [56]) accounts for approximately 50% of clinically severe β-thalassaemia cases [55]. More rarely, the homozygous state leads to the phenotype of thalassaemia intermedia (TI), which is characterized by milder anaemia. Intermedia genotypes mainly include homozygotes (β^+/β^+) or compound heterozygotes (β^+/β^0) for β-thalassaemia and have been characterized by marked phenotypic heterogeneity [56, 57].

The clinical variability of β-thalassaemia in cases of identical primary mutations was explained by identification of important genetic modifiers. First, co-inheritance of α-thalassaemia was shown to confer a milder phenotype, by reducing the amount of excess HBA chains [58]. Second, co-existing mutations activating fetal *HBG* expression in adults alleviate the phenotype, since the HBG chains sequester excess HBA chains into functional HbF [58]. The inherited persistent production of HbF in adults, termed hereditary persistence of fetal haemoglobin (HPFH), has been classified according to the distribution of HbF in erythrocytes, the so-called F cells, into pancellular (pHPFH; homogenous HbF distribution) and heterocellular (hHPFH; heterogenous HbF distribution) forms. The inheritance of the rare pHPFH form follows a Mendelian pattern and is caused by deletions of variable extents of the *HBB* locus or by point mutations in the *HBG1* or *HBG2* promoter. Heterozygosity leads to an increased HbF in the range of 10%–40%, indicating that other genetic determinants influence HbF variability. However, hHPFH does not follow Mendelian patterns and exhibits lower HbF levels. The F cell count and HbF are closely correlated traits. The majority of normal adults have HbF levels <0.6% of the total Hb, whilst only an estimated 10% would carry F cells with HbF levels in the range of 0.8%–5% [59, 60]. Twin studies demonstrated that HbF is a trait with a strong genetic component which accounts for 89% of the quantitative HbF (and F cell) variance [59]. Initial genetic studies identified three loci controlling HbF levels; *Xmn*I-HBG$_\gamma$ in the *HBB* cluster (11p15), *HBS1L-MYB* (6q23), and *BCL11A* (2p16) [60]. Their combined effect on HbF variation was <50%, implicating additional loci, such as the recently identified *KLF1* [61] and *SOX6* [62, 63] (**section 1.4.4**).

1.3.3 Clinical manifestation and pathophysiology

In thalassaemias, the stoichiometry of the main adult haemoglobin HbA ($\alpha_2\beta_2$) is affected by absence or severely lowered expression of one of the constituting globin chains. More specifically, β-thalassaemias are caused by reduced or abolished synthesis of HBB chains, causing excess HBA chains to precipitate, oxidation of released α-haem irons, disturbed membrane stability, and accelerated apoptosis of the damaged RBC precursors, resulting in haemolysis and ineffective erythropoiesis. The ensuing profound anaemia causes expansion of

erythropoietic BM activity, bone deformities, massive extramedullary haematopoiesis and splenomegaly [56] (**Figure 1.6**).

Iron overload is a major and unavoidable complication of β-thalassaemia. Erythroid BM expansion inevitably leads to the accumulation of iron in the body because of inadequate excretory mechanisms. As iron loading progresses, the capacity of serum transferrin, the main iron transporter, to bind and detoxify iron becomes lower. Free iron species, such as labile plasma iron, induces the production of reactive oxygen species, which in turn damages cells causing organ dysfunction and eventually death [53]. Iron loading in the heart presents a major cause of life-threatening morbidity with 6.8% and 5.7% of patients exhibiting cardiac failure and severe arrhythmias, respectively [58]. Also, iron-induced damage to the pituitary gland causes hypogonadotropic hypogonadism characterized by delayed sexual maturation [53], a condition which can be averted if adherence to therapy is maintained within the first decade of life [64]. Iron-induced liver disease, aggravated by viral infection with hepatitis B and C, affects approximately 70% of patients [58], and is usually characterized by cirrhosis within the first decade of life [53].

1.3.4 Management and treatment

1.3.4.1 Prevention programmes

According to the World Health Organization (WHO) records, an estimated of 275,000 children with SCD and 56,000 babies with β-thalassaemia are born each year [65]. In order to manage the health problem of haemoglobinopathies, WHO suggested the development of control programmes that integrate treatment with carrier screening, genetic counselling, and prenatal diagnosis [65]. A successful example includes the thalassaemia-control programme in Cyprus operating since 1973 [66] and achieving an 80%–100% prevention of affected cases per annum.

Population screening was successfully implemented by having all couples provide certificates for their globin gene allele status before marriage. Additional measures of genetic counselling and prenatal diagnosis led to reduction of the annual births of homozygotes to <2% by 2006

[67]. Another successful thalassaemia-control programme has been operated in Sardinia of Italy. A carrier-screening programme in mid-1990s identified 87% of at-risk couples which led to prevention of approximately 90% of affected births via prenatal diagnosis and selective abortion [68].

Control and management strategies are tailored at the country level according to social and religious attitudes, costs, and opportunities within the health system [65, 69]. The successful development of an effective programme entails public awareness, the establishment of laboratory facilities equipped to provide services for disease management, the training of clinicians from the same ethnic or linguistic background to develop expertise in disease management and counselling, carrier screening and prenatal diagnosis [54, 65]. As the proposed strategies are cost-effective compared with life-long management options, even more countries are eager to integrate them in their primary health care system [65]. For example, Iran (1991) implemented the option of aborting affected foetuses in the first 100 days of pregnancy, while Bahrain (1997) and Palestine (2000) enforced the premarital screening and counselling by law [69]. In spite of an increasing cascade of available screening programmes worldwide [69], still, thalassaemia remains a major health problem. In many countries of the developing world, the effective development of these programmes is restricted by major economic and organizational difficulties while many services are inadmissible within societies due to religious and other ethical issues involved [54], ultimately resulting in many children dying undiagnosed, untreated or under-treated. In fact, WHO worldwide estimates reported that approximately 3.4% of deaths in children under the age of 5 are accounted for by severe haemoglobinopathies [65].

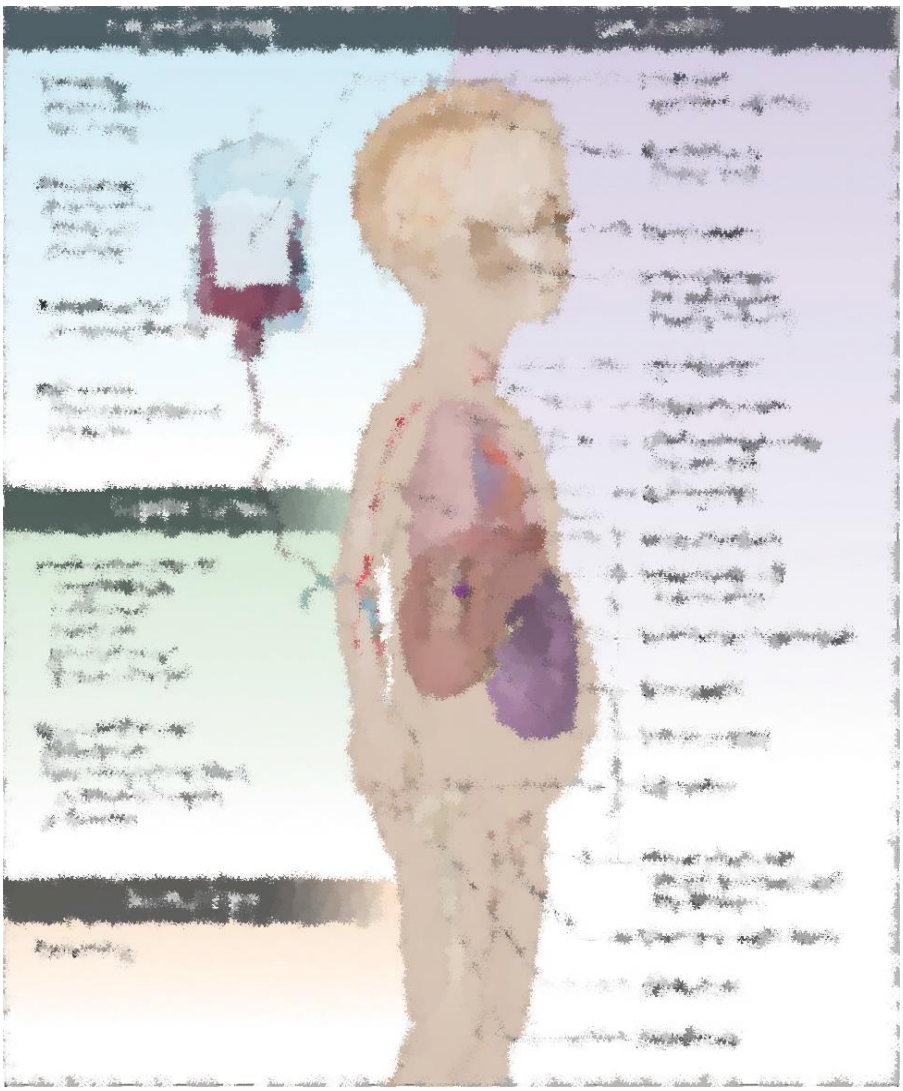


Figure 1.6. Clinical consequences and management options for β-thalassaemia. Reproduced from [70]

1.3.4.2 Blood transfusion and chelation therapy

Transfusion therapy with packed RBCs as the first measure of β-thalassaemia management significantly reduces anaemia without, however, ameliorating toxic iron accumulation, which can be fatal within 10–20 years if poorly managed. Children with TM, if left untreated, die in the first decade of life and most by the age of 5 years. Transfusion therapy aims to increase Hb levels to 130–140 g/L post-transfusion (vs <70 g/L in affected children [56]) and to maintain Hb levels >90–100 g/L [55]. In order to be successful, a life-long treatment with regular transfusion is required, which, however, will lead to iron overload by the age of 10 years [56]. Children with TI

present with mild anaemia later in life (2–6 years of age [55]) and transfusion therapy is tailored to the individual patient's needs, according to the residual level of HbA. Also, children with TI are at risk to develop iron overload as a result of increased gastrointestinal iron absorption because of ineffective erythropoiesis. Efficient though often burdensome iron chelation therapy has helped to extend the life of people with thalassaemia into their fourth to fifth decades of life [64], who nevertheless still have a reduced life expectancy and quality of life compared to healthy individuals.

Iron overload can be effectively treated with various chelating agents. To date, there are three different chelating drugs available, namely, deferoxamine, deferiprone, and deferasirox. All three drugs are capable of chelating non-transferrin-bound iron and promoting its excretion (**Table 1.1**). Deferoxamine was the first drug introduced into clinical practice. Personalised therapy depending on disease severity successfully reduced morbidity from organ system toxicity by reversing hepatic and cardiac iron overload. Although effective, it is an intensive chelation therapy based on the subcutaneous parenteral administration of the drug which, however, presents a rather challenging regime that can frequently result in non-compliance. Deferiprone and deferasirox are more recently introduced orally administered drugs possessing different pharmacokinetics. Deferiprone exhibited significant efficiency in lowering iron levels only in patients with severe iron overload and reported greater efficacy than deferoxamine in removing cardiac iron and improving cardiac function. A combination deferiprone/deferoxamine therapy was suggested for patients with marginal responses to deferiprone. The combination therapy was reported to produce greater efficacy most likely by accessing different iron pools [58, 71]. Deferiprone therapy has been associated with severe agranulocytosis (severe leukopenia; low white blood cell (WBC) count) and arthralgias (joint pain), which are both reversible upon discontinuation of the drug. Other symptoms include nausea, vomiting and abdominal pain, which are mild and can be treated with anti-inflammatory drugs or adjusting the chelating drug dose [58, 71]. Deferasirox is the most recent oral chelating drug and shares deferiprone's capacity to enter cells and bind intracellular iron. Deferasirox reported similar efficacy to deferoxamine in reducing overall iron burden, and was dose-dependent according to the severity of the condition and iron loading. It addition, deferasirox reported efficacy in

reducing cardiac iron as well as preventing iron accumulation in the heart. Deferasirox therapy has been associated with acute renal insufficiency which rarely poses a severe thread and can be resolved with drug discontinuation. Least concerning side-effects include nausea, vomiting and abdominal pain, which are easy to treat [58, 71].

Splenectomy is offered in patients with annual transfusion requirement >200 mL/kg of packed RBCs in order to maintain Hb ≥100 g/dL [55]. It is mainly offered to patients with delayed development and poor health in order to prevent extramedullary haematopoiesis by improving the Hb level, reduce the transfusion requirement and therefore decrease iron accumulation in tissues. The therapeutic benefits are counteracted by an increased risk for postoperative sepsis and thrombosis [55, 72]. Hypersplenism, and therefore splenectomy, can be averted if patients are put early on a comprehensive transfusion and iron-chelation therapy.

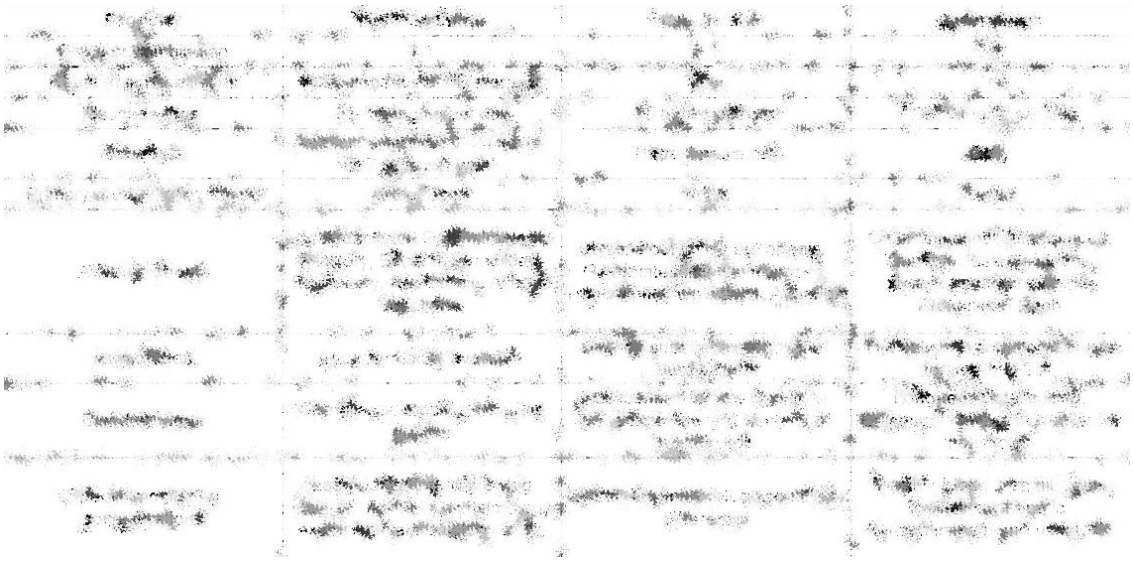


Table 1.1. Comparison of iron chelating drugs. Reproduced from [73].

1.3.4.3 Stem cell transplantation

At present, the only potentially curative treatment for β-thalassaemia is allogeneic bone marrow transplantation (BMT) from a human leukocyte antigen (HLA)-matched donor, preferably a sibling. In 1982 Thomas and colleagues [74] were first to report the treatment of a child who received allogeneic BMT for β-thalassaemia. Since then, it is estimated that >3,000 β-

thalassaemia patients have been successfully transplanted [64]. Paediatric patients who receive regular transfusion and chelation therapy before transplantation report higher disease-free survival than the adult patients (80%–90% vs. <70%) [55]. The success of BMT in paediatric patients can be predicted using the Lucarelli classification which assesses risk factors influencing the outcome and which include the degree of hepatomegaly, the presence of portal fibrosis, and the effectiveness of iron chelation therapy. Low risk patients have good prediction of becoming transfusion independent [75], while high-risk patients with extensive liver damage from iron overload have poor prediction with approximately 30% graft rejection [70]. Adult patients do not conform to the Lucarelli classification scheme since they are usually in a progressive pathophysiological state and heavily iron-loaded. Persistent iron excess after transplantation is usually managed by regular phlebotomy [55, 75]. Other potential complications include graft rejection (≤10%) [75], chronic graft-versus-host disease (GvHD; 2%–8%) [53], and organ dysfunction (e.g., retarded growth and/or hypogonadism) induced by the chemotherapeutic preparative conditioning in addition to iron-overload [64]. Lastly, allogeneic BMT is available only to a minority of ~30% of patients [75].

An alternative therapy for β-thalassaemia patients who do not have an HLA-matched family donor uses marrow graft from an HLA-compatible, unrelated donor. In 2002 La Nasa and colleagues [76] were first to report the treatment of 22 β-thalassaemia patients with allogeneic BMT from unrelated donors. Reported complications included chronic GvHD (25%) and mortality (n=6) mainly secondary to severe liver damage and iron overload. Clinical studies reported an overall survival and thalassaemia-free survival of 97% and 80%, respectively, in low-risk groups, provided that stringent selection criteria for donor compatibility were employed. Unfortunately, predictions for high-risk groups were less encouraging, producing an overall survival and thalassaemia-free survival of 65% and 54%, respectively, mainly in consequence to grade II to IV GvDH or graft failure [64, 75, 77]. In addition, allogeneic BMT using umbilical cord blood (CB) presents a promising alternative and further increases the donor pool. The main limitation of this therapy is the small number of stem cells per umbilical CB sample relative to the number required for engraftment to sustain haematopoiesis and prevent graft rejection [55].

Even though therapy using allogeneic BMT has grown into a feasible curative option for all patients with severe β-thalassaemia, to date, only an extremely small fraction of the affected population worldwide has been treated. Therapy is both intensive and expensive which makes it less accessible in many countries of the developing world where thalassaemia is endemic and most prevalent [54, 78]. In addition, supportive health care has evolved, turning a serious and often fatal disease into a chronic survivable disease with estimated survival into the 6^{th}–7^{th} decades of life [55]. As treatment with HLA-matched grafts bears many risks limiting its success and which is even lower in adult patients who have already developed the clinical complications of thalassaemia, in practice, most eligible patients are hesitant to take up the option, creating a pressing need for new, universal and efficient therapies.

Gene therapy (GT) for the haemoglobinopathies including thalassaemia offers such a more universal curative treatment option. As this entails genetic correction of the patients own HSC problems of allogeneic BM or CB donor availability and complications of graft rejection and GvHD are avoided. Before going on to describe the current status of the field of GT for the haemoglobinopathies, a review of the regulation of expression of the globin gene families is needed to appreciate the progress made.

1.4 Globin gene regulation

1.4.1 Haemoglobin switching

Hb synthesis is controlled by two multigene loci with genes arranged 5' to 3' in the order of their developmental expression to produce different Hb tetramers, a process called 'haemoglobin switching'. The HBA-like gene cluster (5'-*HBZ2-ΨHBZ1-ΨHBA2-ΨHBA1-HBA2-HBA1-θ*-3') is located close to the telomere of the short arm of chromosome 16 in a region rich in CpG islands and ubiquitously expressed genes [79, 80]. The HBB-like gene cluster (5'-*HBE-HBG2-HBG1-ΨHBB1-HBD-HBB*-3') is located in the short arm of chromosome 11 within one of many olfactory receptor gene arrays, which are not expressed in erythroid cells [81, 82]. Each of the globin genes has their own promoter, consisting of DNA motifs (e.g., *cis* elements CCAAT, CACCC and TATA boxes), which bind to complexes of proteins effecting the initiation or silencing of transcription [82, 83]. The erythroid and developmental stage-specific expression of both the *HBA* and *HBB* gene family loci has been attributed to the presence of a strong distal enhancer-type regulatory element. Importantly, in the case of the *HBB* locus, the distal regulatory element retains the ability to override the repressive chromatin environment and activate globin genes sequentially [82]. Co-regulation of the two loci is required for balanced globin chain production in order to form functional tetramers, otherwise anaemia will result [84]. In order to elucidate the regulatory pathways of 'haemoglobin switching' and understand the molecular cause of anaemia, the structure of the mammalian *HBA* and *HBB* loci has been extensively studied, especially in the human and mouse systems.

The human *HBA* locus undergoes one switch from embryonic (*HBZ*) to fetal/adult HBA-like chains (*HBA2, HBA1*) [85] whereas, the human *HBB* locus undergoes two switches, one from embryonic (*HBE*) to fetal (*HBG1, HBG2*) and a second from fetal to adult HBB-like chains (*HBD, HBB*) [84]. During early erythroid development, primitive erythrocytes emerging in the extraembryonic yolk sac are characterized by the production of the embryonic Hb (HbE) Gower 1 ($\zeta_2\varepsilon_2$) and the semi-embryonic Hb types Gower 2 ($\alpha_2\varepsilon_2$) and Portland ($\zeta_2\gamma_2$). With the onset of definitive erythropoiesis in the fetal liver at 5–6 weeks of gestation, the expression of the two *HBG* genes increases to produce fetal Hb (HbF; $\alpha_2\gamma_2$) which is composed of 75% HBG2 and

25% HBG1 chains [86]. The *HBG* genes are the result of a 5 kb tandem duplication and differ at position 136 of the HBG chain to encode glutamic acid ($^{G}\gamma$; HBG2) or alanine ($^{A}\gamma$; HBG1) [87]. Around the 5th month of gestation, erythropoiesis initiates in the BM and HBG chains are gradually replaced by adult HBB-like chains to produce the adult Hbs referred to as HbA1 ($\alpha_2\beta_2$) and HbA2 ($\alpha_2\delta_2$), which account for ~97% and ~2% of the total Hb in adult erythrocytes, respectively. HbF levels decrease to <1% of the total globin expression in the first year of life and it is composed of 40% HBG2 and 60% HBG1 chains [86], co-expressed in a small proportion of adult erythrocytes termed F cells [84, 88] (**Figure 1.7**).

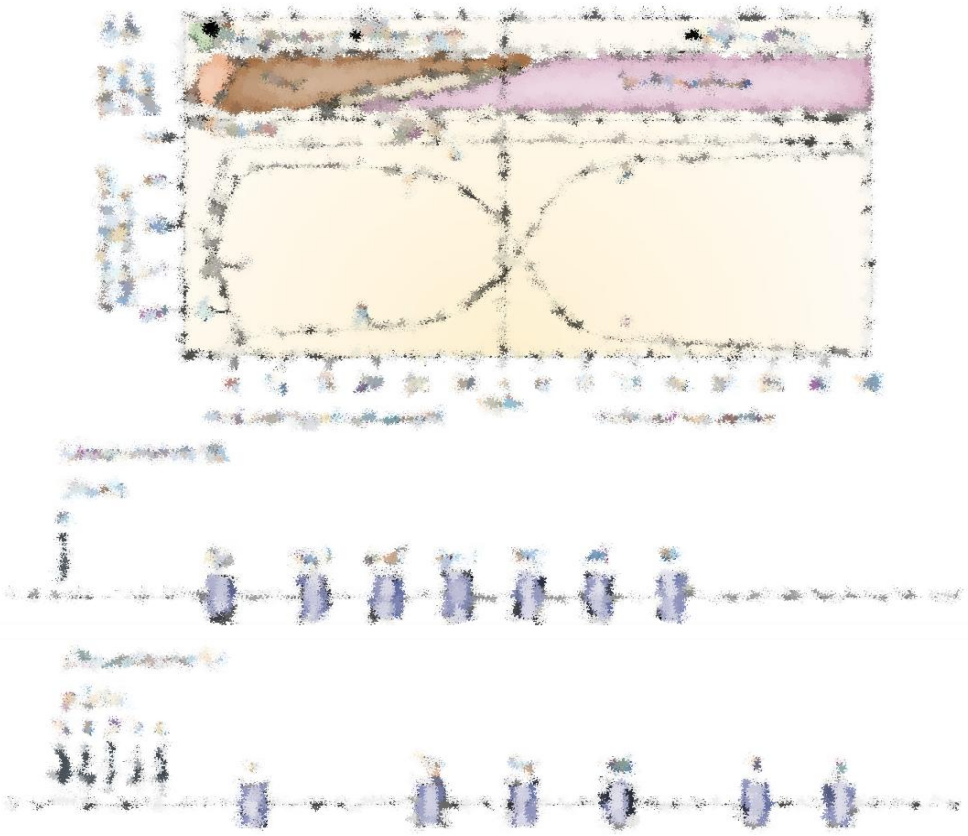


Figure 1.7. Developmental globin switching in humans. The diagram shows the α-like (*HBA*) and β-like (*HBB*) globin gene expression in different sites of human erythropoiesis at various gestational ages. Below, the schematics show the structure of the α-globin locus (chromosome 16) and β-globin locus (chromosome 11); LCR: locus control region, HS: hypersensitive site, Ψ: psudogene. Reproduced from [89].

1.4.2 The *HBA* locus

The duplicated *HBA* genes are in constitutively open chromatin, both being transcriptionally active and encoding identical peptide chains. The *HBA2* gene dominates during development to produce thrice as much HBA chains as the *HBA1* gene, likely owing to differences in promoter sequences regulating transcription [79]. In addition, there are four pseudogenes of which two (*ΨHBA2, HBT1*) retain the transcriptional capacity to produce small amounts of RNA without, however, contributing to the production of functional Hb tetramers [79, 82]. The *HBA* genes are regulated by upstream multispecies conserved sequences corresponding to erythroid-specific DNase I hypersensitive sites (HS-48, HS-40, HS-33, HS-10). Of these, the HS-40 element acts as the major erythroid-specific transcriptional enhancer required to achieve high-level *HBA* expression [90, 91]. Deletions abrogating HS-40 activity markedly reduce *HBA* expression and cause thalassaemia [90]. Essentially, the transcriptional enhancing activity of the *HBA* locus is confined to a single element of ~350 bp, in contrast to the enhancing activity of the *HBB* locus (see below) [90]. The human and murine *HBA* loci and their developmental regulation are similar [80, 92]. Studies in transgenic mice have showed that HS-40 would not confer position-independent expression of the linked transgenes since it does not possess a dominant chromatin opening function, possibly an evolutionary development in response to its natural open chromosomal environment [91, 93]. In consequence, the *HBA*-like genes are expressed in erythroid and non-erythroid cell lines albeit at extremely low levels in the latter. Specifically, the *HBA* promoters do not possess the CCACC and GATA1 sites for binding erythroid-specific transcription factors to fully regulate tissue-specific expression. Critically, HS-40 confers erythroid-lineage specific expression of the linked genes, although, additional (as yet, unknown) regulatory elements present within the wider chromosomal context of the cluster are implicated in tissue-specific control and maximal gene expression [90, 91].

1.4.3 The *HBB* locus

As in all mammalian globin loci, the ordered expression of *HBB*-like genes during development is regulated by a complex transcriptional enhancer-like complex called the locus control region (βLCR). The βLCR consists of 5 DNase I hypersensitive sites (3'-HS5, HS4, HS3, HS2 HS1)

located between 6 to 22 kb upstream of *HBE* [94]. The most potent transcriptional enhancing activity of the βLCR resides in the HS2, HS3, and HS4 elements within 200–300bp core sequences, which constitute an array of binding sites for ubiquitous and erythroid-specific *trans* factors [95]. Highly conserved DNA sequence motifs include the MAF recognition elements (e.g.: NF-E2, NRF1, NRF2, BACH1) and GATA sequences in HS2–HS4, the KLF-binding motif CACCC in HS2–HS3, and an E-box motif in HS2 [94]. Importantly, βLCR-bound proteins have the ability to recruit and interact with macromolecular complexes involved in nucleosome remodelling and histone modifications for the regulation of stage-specific gene expression [94]. Individual HS core elements exhibit specific functions. HS2 behaves as a classical enhancer in the presence or absence of chromatin, and its transcriptional stimulatory activity resides within a conserved DNA sequence motif encompassing a tandem MAF-responsive element (MARE), TGCTGA(C/G)TCA(T/C) [95], of unique configuration to the HS2 site [96]. It is equally active at all developmental stages, contributing substantially to the transcriptional activation of any *HBB*-like gene [97]. In contrast, HS3 and HS4 exert a transcriptional activating function solely within a chromatin context, which clearly distinguishes them from classical enhancer type elements [96]. HS3 functions as an activator of *HBE*, *HBG1 and HBG2* expression during embryonic and fetal stages, respectively, whereas HS4 plays a major role in adult stage expression of *HBB* [97]. HS1 exhibits the weakest transcriptional activity [97] whereas HS5 appears to play a structural role as a chromatin insulator [95]. The βLCR is functionally defined by its ability to confer erythroid-specific, site-of-integration independent, full physiological levels of expression proportional to transgene copy number of a linked *HBB*-like gene. Following on from studies in transgenic mice, it was proposed that the βLCR would overcome repressive position effects at sites of transgene integration to achieve high-level expression at ectopic sites in the genome through an intrinsic dominant chromatin opening and transcriptional activating function [98]. In addition, deletion of the βLCR from the mouse *Hbb* locus, which has similar organization to that in humans [99], did not affect the DNase I sensitivity of the locus, but nevertheless resulted in reduced *Hbb* gene expression (approximately 1%–4% of the wild-type allele [100]). These findings led to the suggestion that the βLCR works through stimulation of transcriptional elongation for the linked genes rather than initiation of transcription *per se*. Furthermore, the

chromatin of the locus remained open in the absence of the βLCR suggesting that other elements, including, but not exclusive, to this element contribute to formation of the open chromatin domain [98, 100]. In humans, the Hispanic β-thalassaemia deletion of the entire βLCR including an additional 25 kb upstream the 5' HS5, is characterized by complete loss of *HBB*-like gene expression. Further analysis of the mutant β-globin gene cluster in a hybrid cell line revealed a closed chromatin conformation and histone hypoacetylation at gene promoters, which rendered the locus transcriptionally inactive [82, 101]. Therefore the large number of factor-binding sites distributed throughout the *HBB* locus, including sequence motifs traversing the deleted region, may function to change its sub-nuclear location and chromatin structure. Currently, many of the *cis*-acting elements harbouring a chromatin opening function remain undefined [95, 101] and could, possibly, explain for differences observed in the chromatin conformation of the murine and human *HBB* loci in the absence of an βLCR.

The transcriptional activation function of the βLCR is tissue-specific. The *HBB* locus remains DNase I resistant in cells, which do not express the genes within this cluster. However, in erythroid cells the locus exhibits a higher level of DNase I sensitivity, which shows variation at more local sub-regions during different developmental stages, reflecting the temporal switch in globin gene expression [94]. In addition, the βLCR is able to enhance expression of linked heterogeneous non-globin gene promoters in erythroid cells. Studies using transgenic assays have shown that the natural function of the linked promoter would confer expression outside the erythroid compartment, suggesting that the tissue-control of basal transcription may reside in the promoter [95]. Deletion studies using transgenic assays identified a functional redundancy between individual HSs suggesting that βLCR elements have specific functions, which are unique and critical for *HBB* expression; e.g., HS3 was able to functionally replace for HS4, whereas HS4 could not replace for HS3 [81, 96]. Current models propose that individual HS elements interact to form a higher-order structure, called the βLCR holocomplex, in which the HS site cores interact to form an active site for *trans* proteins and the core-flanking regions provide the structural framework to constrain the holocomplex in the proper conformation during different developmental stages of erythropoiesis [94, 95]. Deletion of the HS site cores retains an intact holocomplex conformation in which the active site is destroyed, generating a dominant-

negative phenotype that disables the βLCR function. In contrast, deletion of the entire HS site (core and flanking region) enables the remaining HS sites to form an alternative holocomplex conformation with a less effective active site [81, 95]. Incidentally, deletion of individual HSs would result in reduced expression and impaired capacity to protect from position-of-integration effects [96, 97], indicating that an intact βLCR locus is required to achieve full expression of the *HBB*-like genes [81].

The transcriptional activation of globin genes invoked the spatial clustering of *cis* regulatory elements to form a transcription regulatory compartment within the *HBB* locus, called the Active Chromatin Hub (ACH) (**Figure 1.8**). Long-range interactions between the βLCR and the *HBB*-like gene promoters downstream occurred via a DNA/chromatin looping mechanism to form the ACH. Herein, the distal βLCR is brought in spatial proximity to the appropriate globin gene promoter through physical interactions of proteins and co-activators bound to these elements, with intervening DNA (inactive globin and intergenic regions) looping out [81, 95]. The formation of the ACH at the target gene results in high local concentration of transcription factor proteins and chromatin-associated modifiers to establish a nuclear compartment dedicated to RNA-polymerase-II (RNA-pol-II) transcription and thus bring about the required high-levels of *HBB*-like gene expression [102]. The temporal order of chromatin loop formation during development is mainly a function of the changing milieu of factor-binding sites by different globin genes and their interacting ACH *cis* regulatory elements [103]. Some of the factors involved in βLCR/*HBB*-like gene looping are erythroid-specific transcriptional activators (e.g.: KLF, GATA1, NF-E2), co-factors (NLI/Ldb1, BRG1), and insulator-related proteins (e.g.: CTCF, RAD21, SMC1, SMC3) [99, 103]. A parameter relevant to ACH-mediated function is the spacing between the HS elements and the position of the globin genes relative to the βLCR for efficient gene transcription [104]. One particular protein involved in chromatin loop formation in the β-globin locus is the ubiquitously expressed LDB1, which helps facilitate long-range enhancer-promoter interactions and forms a core component of transcription complexes (TAL1, LMO2, E2A, GATA1) that physically link βLCR and *HBB*-like genes during erythropoiesis. The cohesin complex is thought to hold the promoter and enhancer together via formation of cohesion rings around the DNA strands [103, 105, 106]. Another interesting protein influencing chromatin

structural dynamics via the formation of regulatory chromatin loops is the CCCTC-binding factor CTCF. The protein is well-known for its activity at insulator elements in mammalian cells and functions by locating the enhancer and the gene on different chromatin loops, thereby preventing their interactions [99, 103]. The cohesin complex co-localizes with CTCF and appears to help form CTCF-dependent chromatin loops [99].

In addition, the *HBB*-like family are expressed in a competitive manner dependent on their order within the *HBB* locus. This is because only a single gene in the locus can interact with the βLCR holocomplex at any given time to form an ACH [97, 107]. The βLCR will engage and activate the nearest open promoter, and successively form stable interactions with more-distant globin genes as development proceeds [81].

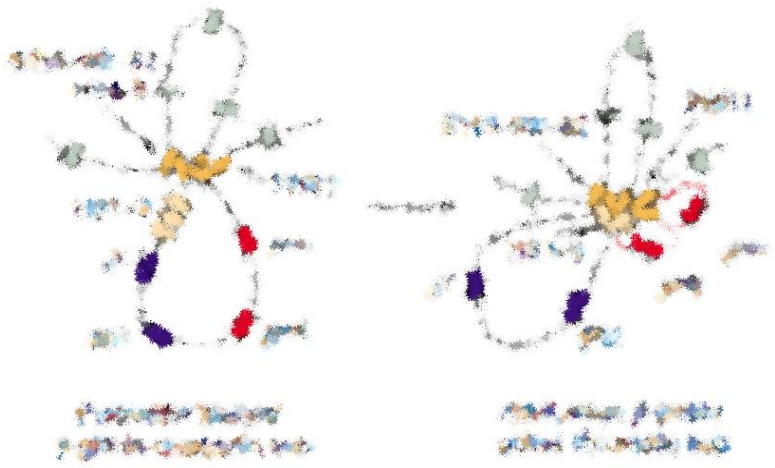


Figure 1.8. Establishment of an active chromatin hub by chromatin looping in the mouse *Hbb* locus. In erythroid progenitor cells not expressing globin, a 'poised' chromatin hub forms consisting of the locus flanking 3' HS1 and 5'HS-60/-62, and the βLCR elements HS4–HS6. In differentiated erythroid cells, the β-globin gene that gets activated and the rest of the βLCR are incorporated to form an active chromatin hub, while the inactive genes loop out. Reproduced from [108].

While the human β-globin locus undergoes two haematopoietic switches, the murine locus performs only one developmental switch from embryonic (*Hbe, Hbbh$_0$, Hbbh$_1$*) to adult chains (*Hbb1*, *Hbb2*) without intermediate fetal step [99]. In transgenic mice harbouring the *HBB* locus, the *HBG* gene is treated as an embryonic gene, and the switch from *HBG* to adult *HBB* occurs during early fetal liver erythropoiesis [109]. Haematopoiesis shifts from the primitive yolk sac

stage to the definitive fetal liver stage at 12 dpc and subsequently to the definitive BM stage at 17.5 dpc from where it is sustained throughout life [109, 110]. Despite similarities in the organization of human and mouse *HBB* clusters, the lack of fetal-specific regulatory elements renders the mouse model problematic in the study of the human *HBB*-like switching.

1.4.4 Molecular factors involved in the fetal to adult switch

The fetal-to-adult globin switch is influenced by a multitude of transcriptional factors and co-activators with binding sites at the gene promoters and/or HS elements of the βLCR, forming an array of large multifunctional complexes with activating or repressing activities. Herein, discussion will focus on BCL11A, KLF1 and SOX6 factors (**Figure 1.9**)**,** which are selected for RNAi-mediated silencing in this study.

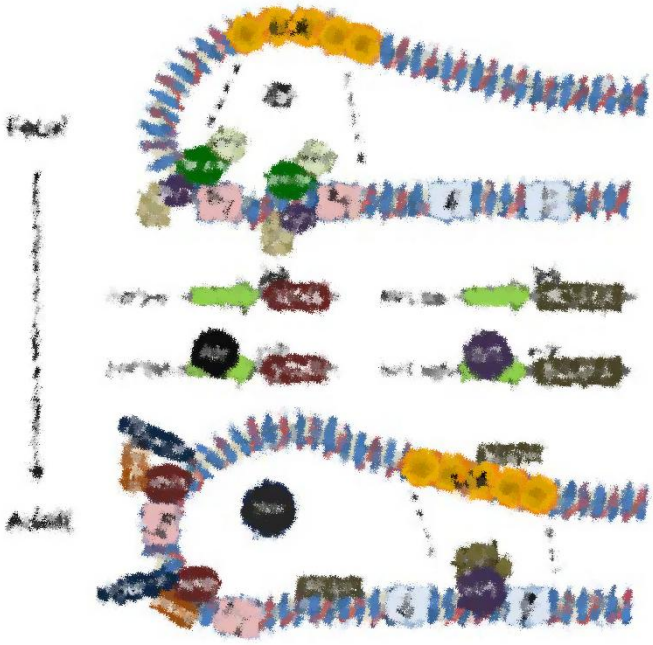


Figure 1.9. Schematic illustration of globin switching model based on long-range chromosomal interactions within the *HBB* locus. Described is how protein-protein interactions may involve chromatin looping to facilitate interaction of distal regulatory elements with cognate promoters and regulate globin gene transcription during development. FOP (Friend of PRMT1) is a chromatin-associated protein modified by arginine methylation and positive regulator of SOX6, with FOP depletion increasing *HBG* expression through reduction of SOX6. NF-E4 is a fetal erythroid-specific transcription factor, which, in conjugation with the ubiquitous transcription factor CP2, facilitates *HBG* transcriptional activation. TR2/TR4/direct repeat erythroid definitive (DRED) complex and COUP-TF (chicken ovalbumin upstream promoter-transcription factor) bind to the direct-repeat site of the *HBG* promoters, silencing *HBG1/2* expression. BCL11A, KLF1 and SOX6 description is provided in the text. Reproduced from [111].

1.4.4.1 BCL11A

BCL11A (B-cell lymphoma/leukemia 11A) is a *Kruppel*-type C2H2 zinc-finger (ZF) transcription factor that was initially identified as a proto-oncogene by virtue of recurrent viral insertions in the mouse orthologue, *Evi9*, resulting in the development of myeloid leukemias or B-cell lymphomas [112, 113]. *Bcl11a*-deficient mice die in the perinatal period from unknown causes whereas, adult mice transplanted with *Bcl11a*$^{-/-}$ cells exhibit impaired B and T cell development, indicating that BCL11A is essential for normal lymphopoiesis [113]. Recent genome-wide association studies (GWAS) have identified a genetic association between the *BCL11A* locus and persistence of HbF expression into adult life [114, 115]. GWAS represent a powerful approach to associate phenotypic variation with allelic heterogeneity by typing a dense array of common genetic markers (e.g., single-nucleotide polymorphisms, SNPs) in a population-based sample that is informative for the trait of interest [116, 117]. The G>A SNP rs4671393 was shown to strongly associate with variation in HbF levels in SCD populations, and the minor allele 'A' was frequently genotyped in patients with high HbF levels and milder SCD [115, 118]. In addition, the 'high-HbF' allele correlated with reduced *BCL11A* expression, suggesting that BCL11A controls silencing of fetal *HBG1* and *HBG2* genes [119]. *BCL11A* expression is developmentally regulated and characterized by the production of four protein isoforms (BCL11A-XL, -L, -S, -XS) throughout ontogeny [120]. Importantly, adult erythroid cells predominately express the full-length BCL11A-XL and –L isoforms whereas, embryonic and fetal cells that highly express HbF, express the shorter BCL11A-S and –XS isoforms [119]. Depletion of BCL11A in human adult erythroid progenitors using short-hairpin RNAs (shRNAs) led to an increase in *HBG* expression without perturbing the progression or transcriptional expression of erythroid differentiation [119]. Therefore it would appear that BCL11A regulates *HBG1* and *HBG2* expression by acting in the *HBB* locus. Indeed, BCL11A occupancy was detected in the HS1–HS3 sites of the distal βLCR, towards the 3' end of *HBE*, and in the intergenic region between the *HBG1* and *HBD* that is specifically deleted in patients with HPFH [62, 63]. Notably, BCL11A did not bind the proximal *HBG* or *HBB* promoters as a mechanism of direct *HBG* repression locally at these sites. Instead, BCL11A induced chromatin reconfiguration of the *HBB* locus in order to facilitate the long-range interaction between the βLCR and *HBG1* and *HGB2* promoters. In transgenic mice

harbouring the *HBB* locus, the βLCR was shown to interact preferentially with *HBB* whereas, upon deletion of *Bcl11a*, the *HBB* locus was reconfigured such that *HBG1* and *HGB2* were placed in a competitive advantage for βLCR interactions and, consequently, suppressed *HBB* expression [63]. In erythroid cells, BCL11A interacts with various partners within cell-specific multiprotein complexes. These include the erythroid transcription factor GATA1 and its cofactor FOG1, the chromatin-associated DNA-binding protein SOX6, and components of the nucleosome remodelling and histone deacetylase (NuRD) complex (**Figure 1.10**) [111, 119, 121, 122]. BCL11A is established as a critical repressor of *HBG1* and *HGB2*, and thus required to maintain HbF silencing in human adult erythroid cells [119].

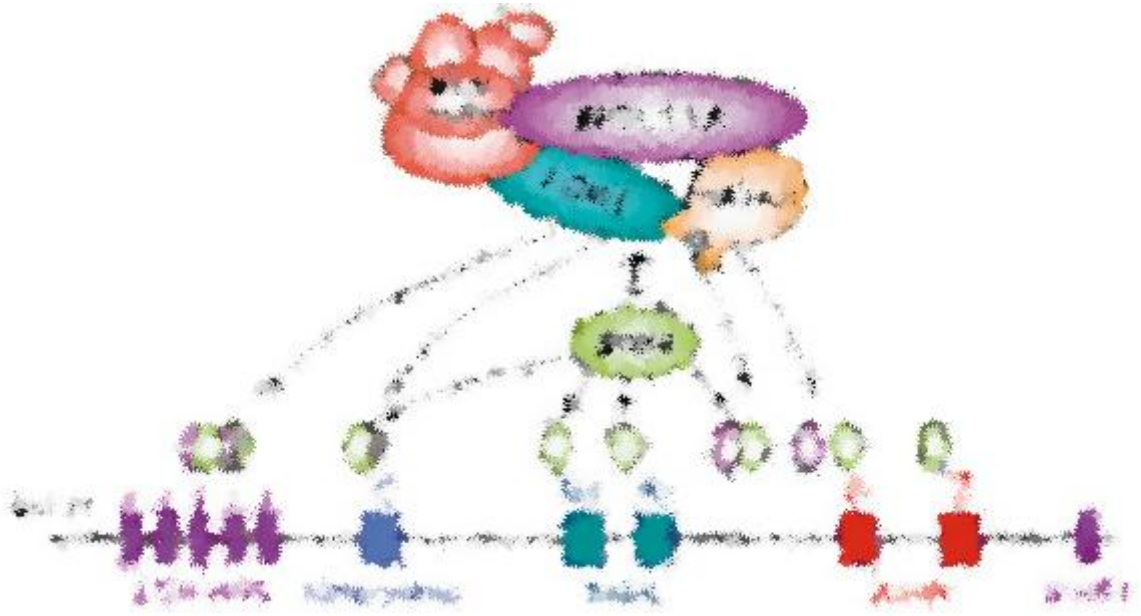


Figure 1.10. Schematic of BCL11A-mediated *HBG* silencing. BCL11A physically interacts with the transcription factors GATA1/FOG1, the HMG-box protein SOX6, and the NuRD remodelling and repressor complex to silence *HBG1/HBG2* in the *HBB* locus. BCL11A does not bind the *HBB* or *HBG* promoters. Instead, it occupies sites within the βLCR and downstream of *HBG1* in the *HBG1-HBD* intergenic region. Reproduced from [63].

As previously mentioned, the expression of *HBB* loci in mice lead to robust expression of *HBG1* and *HGB2* in primitive erythroid cells and their subsequent down-regulation in the definitive erythroid population. Further studies identified BCL11A as a critical mediator of species-divergent developmental globin switching. In this respect, BCL11A exhibits a different stage-

specific expression profile in the mouse compared to that in humans. In the mouse *trans*-acting milieu, BCL11A is expressed in the fetal liver and BM and only as full-length protein isoforms. Knockdown of *BCL11A* expression showed a delay in globin switching and sustained *HBG1* and *HGB2* expression within the definitive erythroid lineage [109]. These observations indicated that BCL11A functions as a developmental regulator of HbF expression and suggested that down-regulation of *BCL11A* or its partner proteins may serve as a promising therapeutic strategy to reactivate HbF in adult patients with β-thalassaemia or SCD. A recent study has demonstrated that the targeted inactivation of *Bcl11a* using the Cre/lox system in *HBB* locus transgenic mice that model SCD, effectively corrected the haematological and pathologic defects associated with the disease through high-level HbF induction [121]. This provided the first proof-of-principle for the therapeutic potential of HbF reactivation in patients with β-haemoglobinopathies. Another study has reported the identification of an erythroid-specific enhancer of *BCL11A*, raising the possibility to achieve an erythroid lineage-restricted deregulation of *BCL11A* with resultant HbF derepression, thus providing an alternative and possibly superior target for therapeutic intervention [123].

1.4.4.2 SOX6

Sox6 is a member to the Sox family of transcription factors with a high-mobility-group (HMG) box DNA-binding domain, acting as regulators of cell fate and differentiation in various lineages [124]. The HMG box domain binds within the minor groove and bends DNA to induce chromatin conformational changes that enhance protein accessibility and plasticity at Sox consensus binding sites, and, thereby, facilitate the formation of multiprotein complexes with activating or repressive activity [125]. The HMG box domain binds preferentially to sequences with a 5'-NNCAAN-3' (N = A or T) motif, and can tolerate mismatches such that putative Sox binding sites can be virtually found in any regulatory region [124, 125]. The Sox family consists of 20 genes. The *SoxD* group – *Sox5, Sox6, Sox13* – has essential regulatory function in the process of chondrogenesis and gliogenesis. *Sox6* alone regulates definitive erythropoiesis [124, 125]. The inactivation of *Sox6* in p^{100H} mutant mice leads to cardiac and skeletal developmental defects, which cause death within two weeks after birth [126]. In addition, a lack of *Sox6* leads

to a persistent expression of embryonic *Hbb*-like globins in definitive erythrocytes. Sox6 is able to interact directly with the *Hbey* promoter and repress expression in definitive erythropoiesis. Intriguingly, Sox6 may associate with other *Hbey*-globin repressors binding near Sox/Sox6 consensus sites, namely the DRED complex and COUP-TF [127], and act in the form of a large repression complex [128]. In addition to its role in globin gene regulation, evidence suggests that Sox6 can engage in the process of red cell maturation [128]. Sox6 is essential for normal erythropoiesis throughout development and has been shown to stimulate erythroid cell survival, proliferation and terminal maturation [129]. Importantly, the level of *SOX6* expression has been proposed to modulate its effect on erythropoiesis [130]. Indeed, cultures of human erythroid progenitor with high-HbF have reduced *SOX6* expression, suggesting that variation in the level of this factor is associated with *HBG* expression [131]. Further work has reported that the interaction between SOX6 and BCL11A, possibly as part of a chromatin-associated multiprotein complex, can facilitate silencing of *HBG* expression in definitive erythropoiesis [63]. SOX6 physically interacts with BCL11A at an exclusive C-terminal three ZF domain. The two transcriptional regulators are co-expressed during erythroid development and co-localise within the human *HBB* cluster. As it also binds the *HBG1 and HBG2* promoters, it has been suggested that SOX6 may help recruit BCL11A and NuRD repressor complexes to the proximal promoter regions of these genes during Hb switching and silence their transcription. Reduction of SOX6 and/or BCL11A activity through shRNA-mediated knockdown in adult erythroid progenitors leads to an increase in HbF (4.7%, 28.9%, and >45% HbF upon knockdown of SOX6, BCL11A, and SOX6+BCL11A, respectively), without any obvious adverse effects on erythroid differentiation *in vitro* [63]. Thus SOX6 knockdown has been considered a promising target for HbF induction. Interestingly, however, a recent report of a patient with a heterozygous disruption of *SOX6* but absence of a concurrent increase in HbF, suggested that expression of this gene may exert a dosage compensation mechanism from the intact allele or that a particular threshold of expression is required to achieve robust HbF induction [132].

1.4.4.3 KLF1

KLF1 (previously named as EKLF) is a member to the *Kruppel* family of transcription factors with a three ZF DNA bidning domain with affinity for the CACCC motif in erythroid-specific regulatory elements including the *HBB* promoter [133]. Mice with homozygous deletion of *Klf1* exhibit normal erythropoiesis at the embryonic yolk sac stage, but they rapidly develop severe anaemia when haematopoiesis switches to the fetal liver. The inactivation of *Klf1* leads to a specific reduction in adult *Hbb* gene expression which causes embryonic lethality [134, 135]. Accordingly, the perturbation of KLF1-binding sites in the *HBB* promoter similarly results in β-thalassaemia [58]. *Klf1*-ablation studies in human *HBB*-locus transgenic mice have shown reduction in *HBB* expression with a concomitant increase in *HBG* expression, suggesting a critical role for *Klf1* in the Hb switch [136]. The *HBB* transcriptional deficiency was associated with reduced DNase I sensitivity at this gene's promoter as well as in the βLCR element HS3, indicating that KLF1 exerts its function through chromatin reorganization [136]. In fact, KLF1 has been shown to interact with the chromatin remodeling SWI/SNF-related complex to promote opening of the chromatin structure and transcription of *HBB* [94, 137]. In addition, KLF1 post-transcriptional modifications can lead to alternative protein interactions that promote transcriptional repression [138]. Klf1 expression has been detected in murine haematopoietic organs at all developmental stages, as well as being largely restricted to erythroid cell lines *in vitro* [133]. KLF1 functions by coordinating the transcriptional activation of a diverse set of erythroid-specific genes involved in the regulation of Hb production, membrane and cytoskeletal integrity, red cell metabolism, and the cell cycle [137].

A number of *KLF1* mutations have been associated with an HPFH phenotype, albeit some have shown a concomitant disruption of erythropoiesis, suggesting that the controlled reduction of KLF1 activity could re-activate *HBG* expression and, therefore, provide a possible therapeutic target [61, 139-141]. The investigation of a Maltese family with HPFH identified a causative mutation (K288X) in *KLF1* which ablated its DNA-binding domain. Importantly, high-HbF individuals with *KLF1* haploinsufficiency exhibited reduced *BCL11A* expression [61]. As the BCL11A promoter is daubed with CACCC sequences, it was proposed that KLF1 can physically

interact with BCL11A to co-ordinate the transcription of the *HBG* genes. Reduction of KLF1 activity through shRNA-mediated knockdown in adult erythroid progenitors led to enhanced *HBG* expression with a concurrent loss of *BCL11A* expression [61, 142]. This was in accordance to events recorded during fetal development whereby low KLF1 levels are insufficient to activate *BCL11A* expression in order to repress the *HBG* genes [142]. In a parallel study, *HBB*-locus transgenic mice carrying a hypomorphic *Klf1* allele exhibited increased expression of mouse embryonic globins and human HBG polypeptides with reduced levels of *Bcl11a* in definitive erythrocytes [142]. A KLF1-BCL11A axis in globin switching has also been described in compound heterozygous Klf1::Bcl11a mutant transgenic mice with repression of *HBG* expression [143]. Therefore KLF1 appears to act as a dual regulator of the fetal-to-adult Hb switch. First, it has been shown to activate *HBB* by interacting directly with regulatory elements in this gene's promoter and secondly, to suppress *HBG* genes indirectly by interacting with the fetal repressor BCL11A [61, 142]. More recently, KLF1 has been shown to occupy regulatory sequences within the MYB-HBS1L intergenic region and partake in the regulation of MYB transcription [144]. The depletion of *Myb* expression in a transgenic adult murine erythroid environment has been found to be associated with decreased expression of *Klf1* and *Bcl11a*, and an increased expression of *HBG*, suggesting that MYB acts as an indirect upstream regulator of BCL11A through KLF1 transactivation [144, 145].

1.5 Experimental therapies

1.5.1 Pharmacological re-activation of HbF

A major goal of experimental therapy for the treatment of β-thalassaemia has been the identification of pharmacologic compounds that induce HbF production in adult patients. To date, more than 70 HbF-inducing agents have been described, the majority of which do not have a verified molecular mechanism and targets. In addition, none of these agents exhibits a combination of efficacy, safety and convenience to use for development to human therapy [146]. Compounds such as 5-azacytidine (5-Aza-C), hydroxyurea (HU), and butyrate analogues have been used most frequently and summarized below.

Early studies have shown that the DNA methyltransferase inhibitor 5-Aza-C has the ability to trigger *HBG* re-activation and increase HbF production to over 20% of total Hb in patients with SCD and β-thalassaemia [147]. In spite of a significant clinical benefit, 5-Aza-C was not pursued further for human treatment over concerns of its potential carcinogenicity. Decitabine is a formulation of 5-Aza-C showing reduced carcinogenic potential and ability to stimulate HbF synthesis at low doses. The mechanism of action has not, as yet, been elucidated, although it is proposed to involve induced hypomethylation at the *HBG* promoters. The drug formulation requires continuous IV infusion over 1–3 hours per day over several days each month, which can result in poor adherence to therapy, while the long-term side-effects are not known. Since 5-Aza-C is a member of the class of cell cycle-specific cytotoxic drugs, it was also speculated that drug treatment could trigger erythroid regeneration secondary to its cytotoxic effects on cycling erythroid cells [146, 148]. This notion further prompted the evaluation of other cytotoxic compounds, including HU.

HU has been formulated as an antineoplastic drug (e.g., against polycythemia vera), which halts DNA synthesis via inhibition of the enzyme ribonucleotide reductase and arrests cells in S phase. It is taken orally, is relatively inexpensive and has shown a good record of safety for long-term administration. Initial studies for the treatment of SCD have shown induction of HbF with reduction in neutrophils and reticulocytes, and improved RBC survival with no evidence of

toxicity. Clinical findings have shown that HU reduces the frequency of painful crisis and acute chest pain in SCD patients, and can reduce their RBC transfusion requirements. Even though HU does not seem to prevent stroke, nonetheless, it was shown to reduce the incidence of secondary stroke in SCD paediatric patients. In addition, HU has proved to be effective in approximately 60% of SCD patients and in a lower proportion of β-thalassaemia patients. Responders to HU treatment show variable HbF levels in the range of 4%–20%, more likely in consequence to other underlying genetic determinants. The mechanism of inducing HbF production is independent of ribonucleotide reductase inhibition. It is speculated that HU acts on primitive precursors and triggers HbF production either directly or indirectly, the latter mediated by killing the dividing late erythroid cells. This mechanism remains unproven. In 1998, hydroxyurea was approved by the FDA (US Food and Drug Administration) for the treatment of SCD in adult and paediatric patients [148, 149].

Butyric acid analogues are members to the class of short-chain fatty acids and well-known inhibitors of the family of histone deacetylase enzymes. Their HbF inducing activity was initially identified in children born to diabetic mothers (whose blood contains elevated plasma α-amino butyric acid levels [146, 148]), and validated *in vitro* showing increased HbF synthesis with low toxicity. Arginine butyrate resulted in a significant increase in HbF levels in patients with SCD that was, however, not sustained with continuous therapy. Arginine butyrate has a very short half-life *in vivo* which requires a long and continuous intravenous infusion at the expense of a growth-inhibitory activity on haematopoietic cells. An alternative intermittent administration for 4 days every 4 weeks resulted in sustained HbF production in the range of 2%–20% in the majority of SCD patients with exception in those having <2% HbF baseline levels. Sodium phenylbutyrate and isobutyramide are orally administered compounds shown to result in a significant stimulation of HbF synthesis in patients with SCD and β-thalassaemia. The drug protocol requires the intake of 30–40 pills per day, which can result in poor adherence to therapy. Also, elevated HbF was not sustained with continuous therapy and success was restricted to patients with high pre-treatment HbF levels. The mechanism of action is not known. However, it is thought that butyrates act by allowing certain transcription factors to resume activity at the *HBG* promoter regions associated with acetylated histones [147, 150].

1.5.2 Gene therapy

Gene therapy (GT) based on autologous transplantation of genetically modified stem cells is a promising therapeutic option for patients with haemoglobinopathies. The objective of the therapy is the introduction of nucleic acids into cells that will modify gene expression in order to prevent, halt, or reverse the pathological process [151]. Depending on the basis of the pathogenesis, different GT strategies can be considered [152]:

- Gene augmentation therapy [153]. This approach is best suited for loss-of-function mutations whereby introducing normal copies of the mutated gene may increase the amount of normal gene product to the level required for therapeutic benefit. This is the commonest form of GT.
- Gene knockdown [154]. This approach is useful for autosomal dominant diseases caused by gain-of-function mutations and/or in cases of haploinsufficiency. The mutated allele is supressed, without altering the expression of the normal copy, by targeting the mutant mRNA for its specific destruction using RNAi tools. This approach is also used to inhibit cryptic splice sites, which can lead to expression of an aberrantly spliced protein with deleterious effects.
- Gene correction/replacement [155]. Also called ‘genome editing’, this technology uses chimeric nucleases composed of a sequence-specific DNA binding domain fused to a nonspecific DNA cleavage domain. The DNA binding domain is derived from ZF nucleases or transcription activator-like effector nucleases (TALENs) and can be customized to recognize virtually any sequence. The latest development of these tools uses RNA-guided nucleases based on the CRISPR/Cas system, consisting of a short guide RNA with sequence homology to the target DNA and a nuclease (e.g., the Cas9 endonuclease) for cleavage. The chimeric nucleases induce site-specific DNA double-strand breaks (DSBs) that stimulate the cellular DNA repair mechanisms of non-homologous end joining (NHEJ) and homology-directed repair (HDR). NHEJ directly re-joins the DSBs whereas HDR utilizes a DNA template as a substrate for rejoining. HDR enables replacement of the mutated stretch of DNA with its functional counterpart. This

technology is capable of correcting the underlying cause of the disease, alleviate its symptoms and offer the opportunity to patients for a personalized therapy. Currently, genome editing tools are at the forefront of biomedical research.

There are two strategies of gene transfer for the treatment of hereditary disease; *ex vivo* gene transfer and *in vivo* gene transfer. The *ex vivo* gene transfer involves the removal of target cells from the patient, the administration of the therapeutic nucleic acid into these cells in culture (*in vitro* procedure), and the transplantation of the modified cells back to the patient from which they were derived to establish a stable graft expressing the desired protein. In contrast, *in vivo* gene transfer involves the direct administration of the therapeutic nucleic acid into the patient's body either by direct injection to the required tissue or by intravenous injection into the circulation that feeds the tissue. Depending on the molecular basis of a genetic disorder, some GT strategies are better suited to particular types of disorders [153]. For example, GT for β-thalassaemia is performed using *ex vivo* gene transfer (**Figure 1.11**), because the target cell population (BM or peripheral blood-derived HSCs) can be easily obtained from the patient. In addition, it has also been demonstrated that the genetically modified HSCs can exhibit a selective survival and proliferative advantage *in vivo* in certain cases with potential for long-term correction of the pathology [156].

In order to achieve high-efficiency gene transfer and sustained gene expression for effective therapy, the exogenous genetic material must be delivered to the cell's nucleus. Physical transfer of genes into targeted cell populations can be performed by electroporation, hydrodynamic injection, sonoporation or use of ballistic particles (gene guns). However, these techniques suffer from reduced survival when applied to small and delicate cells such as the HSCs [157]. Moreover, if the genetic material is delivered into the systemic circulation in a naked form, it cannot protect itself against phagocytosis and degradation by serum endonucleases in the blood [158]. Hence, two types of vectors have been developed for gene delivery; viral and non-viral. Viruses have evolved to infect mammalian cells and hijack their cellular machinery for self-replication. These viral properties are attractive prerequisites for an efficient vector system. Therefore for therapeutic application, viral vectors have been

engineered for safety by making them replication incompetent [159], while their capacity for cellular targeting and transduction efficiency has not been altered [160]. Alternatively, a range of non-viral 'synthetic' vectors have been developed based on cationic lipids or polymers, which bind and shield the genetic material for systemic or ectopic (e.g., the lungs) delivery. Different targeting moieties have been included in non-viral vector formulations in order to enhance transfection efficiency. Limitations of these synthetic vectors mainly include low efficacy owing to poor stability and biodegradability, as well as cytotoxicity and overall poor transfection capability [157, 160, 161]. An improvement to this strategy has included the use of inorganic nanoparticles, which can be tailored to acquire versatile properties similar to that of a virus for efficient gene delivery. At present, these synthetic vector designs require further optimization to improve efficiency and specificity [160].

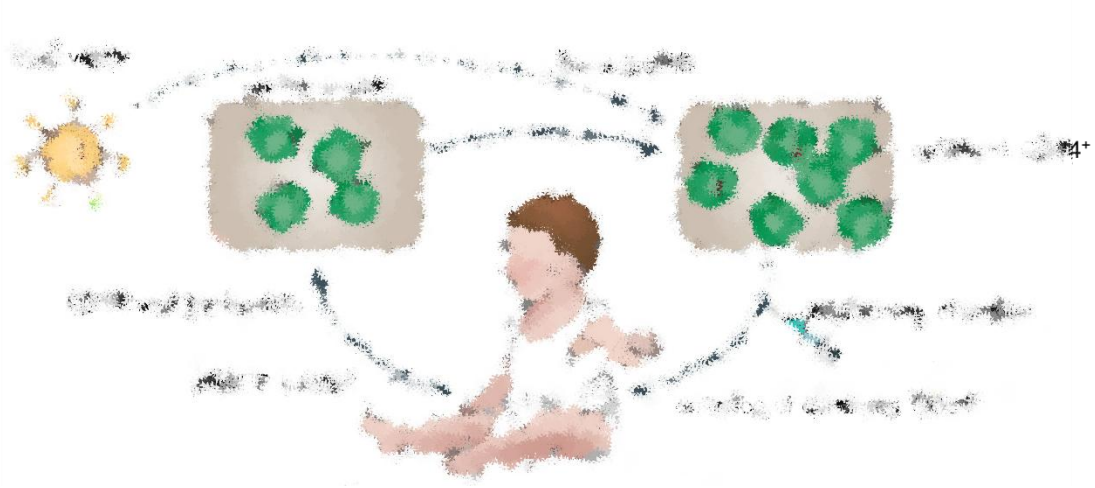


Figure 1.11. E*x vivo* HSC gene therapy procedure in the context of β-thalassaemia. $CD34^+$ HSCs are isolated from BM or peripheral blood following cytokine mobilization of a β-thalassaemia patient, and are incubated *ex vivo* in cytokine-supplemented media that stimulate cell proliferation and enhance cell transducibility with retroviral class of vectors. Stimulated HSCs are transduced with a retroviral-type vector carrying a functional copy of the defective *HBB* gene, followed by quality control. The patient is treated with moderate non-ablative or full myelosuppression, which aims to open BM spaces by eliminating endogenous HSCs in order to facilitate engraftment of genetically corrected HSCs. Modified from [162, 163].

The Journal of Gene Medicine Clinical Trial Database (Wiley Database accessible on-line at http://www.abedia.com/wiley/index.html) contains information on 2075 completed, ongoing or pending human GT clinical trials worldwide over the years 1989–2014. Most GT clinical trials use a viral vector gene delivery system (over 70%) with most derived from retroviruses (23.3%), adenoviruses (22.8%), and adeno-associated viruses (AAV; 5.5%). Different classes of viruses have unique features useful for GT (**Table 1.2**) and the decision for the viral vector best suited for a particular genetic disorder mainly depends on the requirement for either long-term or transient gene expression, as well as the infectivity of the target cell population [164]. Therefore these viruses can be grouped according to whether they integrate into the host genome (retroviruses) or whether they persist in the cell nucleus predominately as extrachromosomal episomes (adenoviruses and AAVs) [165].

At present, integrating vectors based on retroviruses are preferred for therapeutic applications, which require a stable genetic modification in dividing cells [165]. A major breakthrough in GT was achieved ~15 years ago using a gamma retrovirus-based vector to correct $CD34^+$ HSCs for the IL2RG-cytokine receptor in 11 children with fatal X-linked SCID-XI immunodeficiency syndrome. Therapy successfully restored the immune system [166] but long-term clinical benefit was compromised by insertional activation of the *LMO2* proto-oncogene and the development of a leukaemia-like disorder in some cases [167]. A later trial for chronic granulomatous disease, which is caused by a defect in the oxidative antimicrobial activity of phagocytes, produced similar findings. Following initial success, transgene expression was silenced by methylation of the SFFV viral promoter driving expression of the therapeutic (gp91phox) gene and myelodysplasia was developed in consequence to insertional activation of growth-promoting genes [168, 169]. These early studies emphasized the need for the development of vector systems which are highly-efficient and safe.

More recently, research has focused on the use of lentiviruses (specifically, human immunodeficiency virus, HIV), which are also members of the *Retroviridae* family of viruses. Lentiviral-based vectors (LVs) present an alternative therapeutic gene delivery system to gamma retroviruses (specifically, Moloney murine leukemia virus (MLV) [170]). The main

advantage of LVs over MLV-based vectors is their ability to transfer genes in non-dividing cells, such as $CD34^+$ HSCs. (Note: MLV-based vectors can only transduce dividing cells, mainly because they are not equipped to traverse the nuclear membrane) [153]. Although LVs share the risk of gamma retroviruses for insertional mutagenesis, they are considered to be safer because of they lack the propensity for integration in and around promoter regions [171]. To date, LVs have been used in about 4.2% of clinical trials worldwide.

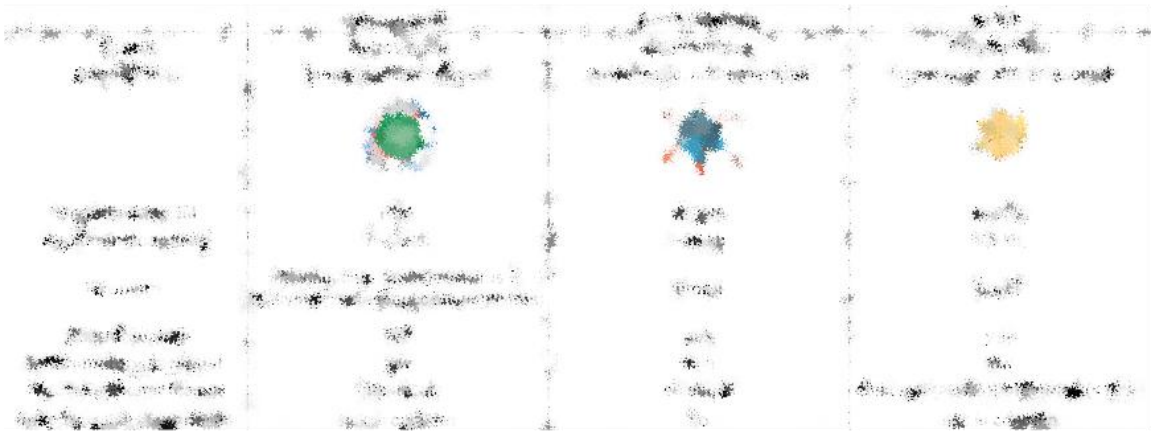


Table 1.2. Features of main gene therapy viral vectors. Adapted from [153, 165, 172].

1.5.2.1 Preclinical studies for β-thalassaemia

Globin-based LVs have become the most effective gene-transfer system for GT in haemoglobinopathies, mainly owing to their ability to stably transduce non-dividing HSCs, package full-length unspliced RNA, and accommodate relatively large (~8kb) transgene cassettes [173]. Modified to express globin genes under optimal regulatory sequences (βLCR components), LVs under preclinical development aim to achieve i) efficient gene transfer and stable transduction of HSCs, ii) controlled transgene expression (erythroid- and stage-specific, elevated, sustained, and position independent), and iii) absent or low genotoxicity [174, 175]. A limitation of early attempts was the requirement for multiple vector copies per cell for phenotypic correction because of variable transgene expression. Notably, the higher the vector copy number (VCN), the greater the risk for insertional mutagenesis. Therefore further optimization to vector design included the insertion of insulator elements, which possess i) chromatin barrier

activity to reduce position effects (e.g., from heterochromatin) silencing transgene expression, and ii) enhancer blocking activity to prevent promoter-vector enhancer interactions at sites of integration leading to undesired, inadvertent host gene activation [176, 177]. Insulator elements can facilitate a sustained vector-derived expression, which could possibly correct the disease phenotype. However, the efficacy of insulated vectors is often compromised by reduced viral titres (e.g., the chicken HS4 insulator, cHS4 [176, 178]), which may hinder large-scale vector production for human trials. Currently, there is an on-going search for alternative insulators with ability to improve efficacy and safety [179]. Recently, the ankyrin insulator element was included in the T9W globin-based LV with minimal effect on titre while also showing high and stable vector-derived expression [180]. Furthermore, a recent proof of principle study demonstrated the possibility of inserting exogenous genetic material into 'safe harbour' sites of the genome using iPS technology, an approach with an unclear long-term safety profile [181]. The HBB expressing vectors currently under preclinical development for GT (**Figure 1.12**) have shown sustained and therapeutic HBB levels correcting the anaemia and secondary organ damage in murine models of thalassaemia.

Success of globin GT was first demonstrated with the TNS9 LV in severe TI mice ($Hbb^{th3/+}$) with an average haemoglobin increase by 3 to 4 g/dL per VCN [182]**.** Haemoglobin levels remained stable in long-term primary transplanted mice, durably correcting anaemia [183]. Using the same vector in a murine TM model system gave variable *HBB* expression, which was sub-therapeutic, converting the TM phenotype (2–4 g/dL) into that of severe TI (4–8 g/dL) [184]. These *in vivo* studies by the Sadelain group were the first to show stable transmission and therapeutic *HBB* expression from within LVs carrying a βLCR HS2/HS3/HS4 plus a mini-*HBB* combination. A similar vector by the Leboulch group, carrying a modified anti-sickling β-globin gene (βT87Q), exhibited lower *HBB* expression and required an average VCN of 3 per cell to stably correct murine TI [185, 186]. Similarly, this vector failed to correct murine TM, demonstrating the need for vectors with higher and more sustained *HBB* expression. In a subsequent study by the Malik group, a cHS4-insulated LV termed BG-I successfully corrected TM in human $CD34^+$ cells, attaining HBB levels similar to those of normal BM progenitors with an average VCN of 2 per cell. This result was confirmed *in vivo* with TM xenograft $\beta2m^{null}$-NOD-

SCID mice [178, 187]. In a recent study by the Rivella group, an ankyrin-insulated LV termed AnkT9W was shown to correct TI in human $CD34^+$ cells, achieving 35% HbA with an average VCN of 0.88 per cell. However, this same vector showed variable HbA expression in the range of 0%–62% in TM human $CD34^+$ cells with a VCN of 0.92 per cell, suggesting that a VCN >1 is required to correct TM in human $CD34^+$ cells [180]. An LV by the Antoniou/Ferrari groups termed MA821 (or GLOBE) and containing only βLCR elements HS2/HS3 plus a mini-*HBB* showed complete and persistent correction of TI in primary and secondary transplant mice ($Hbb^{th3/+}$) with an average VCN of 1 per cell, but with higher vector copies required to correct murine TM [156]. Preclinical analyses in human cells showed correction of TM in $CD34^+$ cells with a VCN of only ~1.6 [188]. The same group, quite recently, showed that the inclusion of the GATA1-HS2 enhancer into GLOBE (re-named to G-GLOBE) successfully corrected TI in primary and secondary transplant mice ($Hbb^{th3/+}$) at a lower VCN than the original GLOBE LV (VCN 1 vs VCN 1.7), while also maintaining elevated transgene expression by protecting the integrated transgene from negative chromatin effects [189].

A novel GT strategy targeting the β-haemoglobinopthies was recently reported where an artificial ZF protein was designed to tether the transcription co-factor Ldb1, which participates in βLCR/*HBB* looping and ACH formation to the proximal *HBG* promoters. In cultures of primary human erythroblasts, the ZF-Ldb1 fusion protein targeting both *HBG* genes, forced their juxtaposition with the βLCR by modulating chromatin structure and stimulated their transcription to approximately 85% of total globin (HBG + HBB) synthesis with a corresponding reduction in adult *HBB* expression. This degree of increased HbF production would possibly alleviate clinical symptoms while, in the context of SCD, the diminished synthesis of defective HBB chains would further improve anti-sickling efficacy [190]. Therefore manipulation of chromatin looping as a means to control globin gene expression represents an appealing strategy to develop therapeutic applications for the treatment of β-haemoglobinopathies.

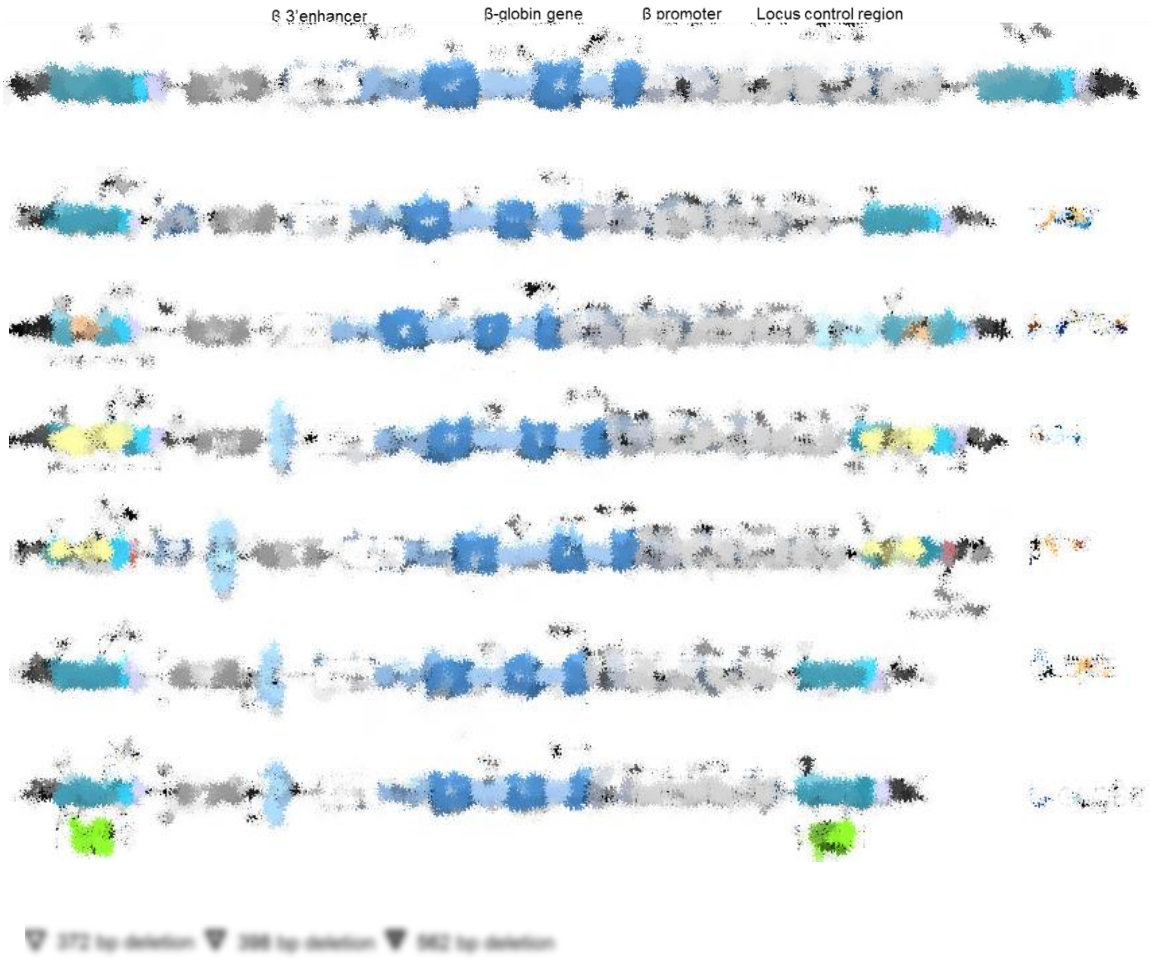


Figure 1.12. Lentiviral vectors expressing HBB. A prototypical design of the HBB lentiviral vectors (LVs) and its main components are shown in the top panel. The schematics of successful LVs used in preclinical studies for β-thalassaemia correction are shown as indicated. Adapted from [191]. HS: hypersensitive site; cHS4: chicken hypersensitive site-4; βp: *HBB* promoter; UTR: untranslated region; 3' enh: *HBB* 3' enhancer element; cPPT: central polypurine tract; SA: splice acceptor site; SD: splice donor site; RRE: REV response element; Ψ: packaging signal; WPRE: woodchuck hepatitis virus posttranscriptional regulatory element.

1.6 RNA interference

RNA interference (RNAi) was initially described as a mechanism of post-transcriptional gene silencing in 1998 in the nematode *Caenorhabditis elegans* by Fire and colleagues, who reported the capacity of exogenous double-stranded RNAs (dsRNAs) to produce specific and robust gene silencing by triggering the cleavage of endogenous mRNA transcripts in a sequence-specific manner [192]. Remarkably, RNAi induces systemic silencing in plants and nematodes, and appears to be a heritable trait, transmitted to the progeny of dsRNA-injected animals even though inheritance beyond the F1 generation is rather rare and only applicable to some germline targeted genes [193]. The RNAi mechanism is conserved across a diverse group of organisms. Several types of regulatory small RNAs have been discovered, including small interfering RNAs (siRNAs), microRNAs (miRNAs), and PIWI-interacting RNA (piRNA) [194]. Different small RNAs use diverse and complex regulatory pathways, which interact at several levels competing for and sharing effector proteins as well as cross-regulating with each other [195]. The exogenous siRNA and endogenous miRNA pathways are converged during biogenesis in spite of their distinct modes of gene regulation, acting predominately via transcript degradation and translation repression, respectively [196]. The distinct class of piRNAs function predominately in the germline and associates with silencing of mobile repetitive elements, such as transposons, which risk insertional deregulation of gene expression [195]. This section will focus on siRNA and miRNA biogenesis and their silencing mechanisms.

Shortly after the discovery of the RNAi phenomenon, which resulted in AZ Fire and CC Mell being awarded the Nobel Prize in Physiology/Medicine [197], dsRNA-based silencing in plants [198] and extracts of *Drosophila* cells [199] demonstrated that the long dsRNA trigger is converted by enzymatic processing into short dsRNA fragments of 20–25nt in length. Thus these short dsRNA species were identified as the effector molecules of dsRNA gene silencing since degraded mRNA was cleaved at sites of sequence homology to the short dsRNA species at 20–25nt intervals. This new class of RNA-based signalling molecules was termed siRNAs [200]. Acting as defenders of genome integrity, siRNAs are produced from foreign sequences themselves, such as viruses, transposons, and aberrant transcription products, co-opting them

into the siRNA mechanism to suppress their expression [194, 201]. Therefore siRNAs are exogenous and have the ability to degrade invasive genes with almost perfect complementarity [202].

In mammals, miRNA-mediated gene regulation is involved in many developmental and cellular processes [197]. The first miRNA, *lin-4*, was described in 1993 in the nematode *C. elegans* by Ambros and colleagues, in a screen for genes involved in post-embryonic development [203]. Particularly, *lin-4* produced a pair of small RNAs sharing antisense complementarity to sequences present in the 3' UTR of the *lin-14* gene. Binding of the miRNA to the 3'UTR target site reduced lin-14 protein synthesis without affecting mRNA levels and resulted in larval developmental arrest, indicating that the *lin-4* miRNA acted to suppress translation elongation or termination rather than translation initiation and proposed a function of miRNAs as regulators of endogenous genes [203, 204]. The miRNAs are genomically encoded and are transcribed by RNA-pol-II as long primary transcripts (pri-miRNAs) [205]. Although some miRNAs are individually transcribed from single units, many more miRNAs are derived from transcription units encoding more than one product. The latter type of transcript includes clusters of distinct miRNAs as well as those located within intronic regions of mRNA encoding genes [204]. Intronic miRNAs are released from the gene transcript (pre-mRNA) by RNA splicing and fed into the miRNA biogenesis pathway [202]. In rare cases of short intronic hairpins, which lack the lower stem of the pri-miRNA and form pre-miRNA-like introns, called mirtrons, the production of pre-miRNAs skips processing by Drosha and depends exclusively on the nuclear pre-mRNA splicing pathway. A lariat-debranching enzyme is involved to produce a hairpin fold suitable for entering the canonical miRNA pathway, indicating that mirtrons is an alternative source of miRNA-type regulatory RNAs [206]. Importantly, miRNAs are endogenous contributing approximately 1% of the total gene content [207] and regulating the expression of at least 30% of human genes [208]. The miRNA precursors fold into imperfect hairpins containing various mismatches, internal loops and bulges and function to suppress translation via a process of partial complementarity [196].

1.6.1 The miRNA and siRNA biogenesis pathways

The miRNAs are processed from long primary RNA-pol-II transcripts into 20–25nt regulatory RNAs via consecutive cleavage steps by two multidomain ribonuclease III (RNase III) endonucleases, Drosha and Dicer. In the nucleus, Drosha assisted by its dsRNA-binding domain (dsRBD) partner protein DGCR8 (in mammals), recognizes and cleaves the pri-miRNA at the base of an internal stem-loop releasing it from the flanking single-stranded arms. The resulting hairpin is excised as a 60–70nt precursor, termed pre-miRNA, containing a loop flanked by base-paired arms and having a 2nt overhang at its 3' end and a phosphate group at its 5' end. The pre-miRNA is exported from the nucleus to the cytoplasm by the RanGTP-powered transporter exportin 5. In the cytoplasm, Dicer assisted by its dsRBD partner protein TRBP (in mammals), recognizes and cleaves the pre-miRNA. Importantly, the PAZ domain of Dicer binds the 3' overhang created by the Drosha/DGCR8 holoenzyme and the two catalytic RNase III domains of Dicer form an intramolecular dimer to create the active site for cleavage of the pre-miRNA to remove the loop. The resulting structure, called miRNA duplex, consists of a 20–25nt dsRNA fragment bearing 2nt 3' overhangs on either strand. In addition to producing miRNAs from pre-miRNAs, Dicer cleaves long dsRNAs to produce siRNA duplexes using the 5' end counting rule [209], which entails measuring a fixed distance from the 5' termini to cleave both strands of RNA. Therefore the miRNA and siRNA biogenesis pathways converge at the Dicer processing stage. Furthermore, Dicer contains an ATPase/RNA helicase domain which unwinds the dsRNA duplex. Essentially, the 'guide' strand will associate with the RNA-induced silencing complex (RISC) whereas the 'passenger' strand will become degraded. The determinant of which strand gets loaded into RISC is based on the relative stability of the si/miRNA duplex termini. The selected strand is usually antisense (complementary to the target gene) with a thermodynamically less stable 5' end relative to the 3' end within the duplex. Loading to the RISC complex is mediated by a rate-limiting unwinding step which allows the 5' end of the strand positioned at the weakly base-paired end of the dsRNA duplex to enter first. The single-stranded si/miRNA associates with Argonaute proteins (AGO1–4) residing the RISC complex and which coordinate processing of homologous mRNA. The mechanism of siRNA-triggered mRNA degradation requires the catalytically active AGO2, which mediates cleavage in

the middle of the complementary region. In contrast, miRNA duplexes consisting of central mismatches preferentially interact with non-catalytic AGO proteins, which guide translational regulation. For this to occur, a small 2–7nt sequence in the 5' end of the miRNA, called the 'seed' region, has to be perfectly complementary to the target mRNA. Because the target specificity is directed through a small complementary sequence, it is possible that individual miRNAs would target more than one mRNA containing the seed binding site. The si/miRNA pathways are illustrated in **Figure 1.13** [195, 205, 210-213].

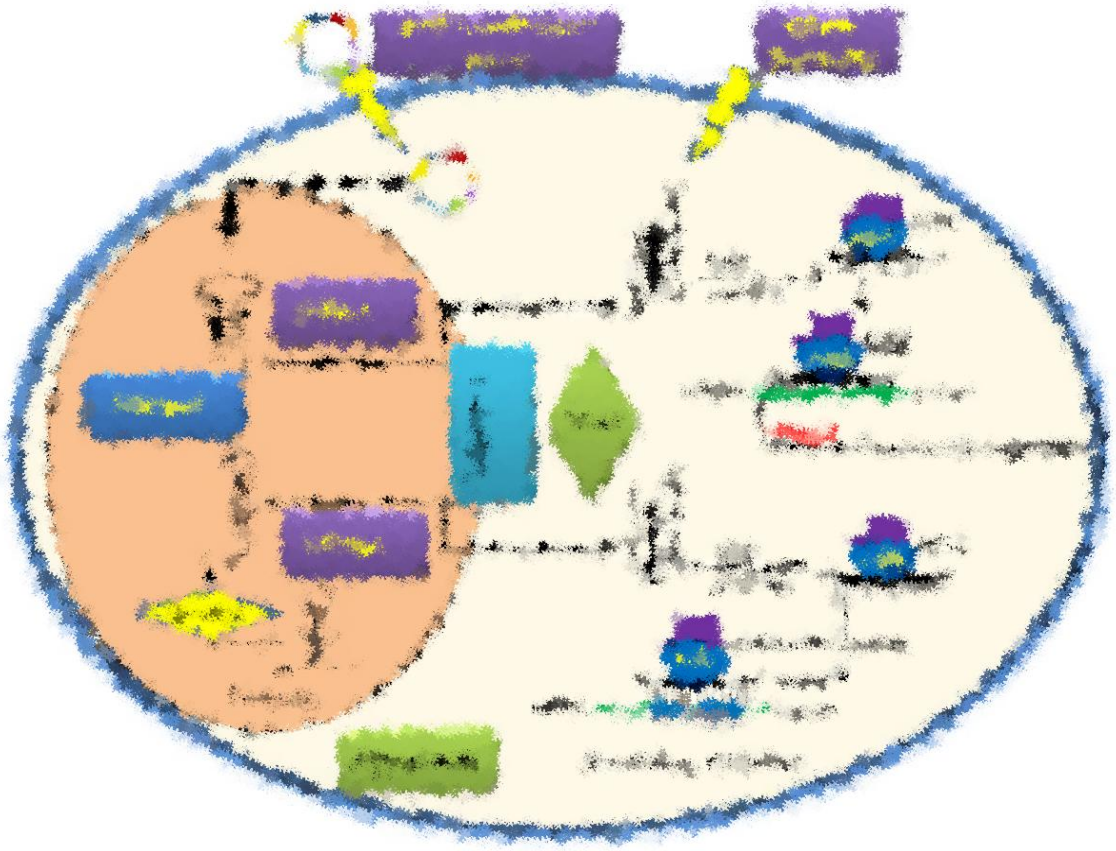


Figure 1.13. Mechanism of miRNA, siRNA, and shRNA biogenesis and mechanism of action in mammalian cells. Reproduced from [212].

1.6.1.1 siRNA-mediated transcriptional gene silencing

The siRNAs with sequence homology to the promoter region of a gene have been shown to trigger gene silencing at the level of transcription, a process termed transcriptional gene silencing. Essentially, siRNAs trigger the epigenetic modification of homologous chromatin at target genomic loci, promoting the formation of heterochromatin resulting in transcriptional repression. These heterochromatic regions include dense H3K9 di-methylation across the homologous DNA. This process is best characterized in the fission yeast *S. pombe* but, as yet, remains poorly understood in the mammalian system. The single-stranded siRNA incorporates into the RNA-induced transcriptional silencing complex (RITSC) by binding the Argonaute protein Ago1, followed by localization to the nascent RNA transcribed by RNA-pol-II. It is speculated that the interaction between Ago1 and RNA-pol-II helps recruit the RITSC to the gene promoter, while chromatin interaction is facilitated by the RITS chromodomain protein Chp1. The RITSC targets the recruitment of epigenetic remodelling complexes which catalyse H3K9 methylation in the target chromatin, creating selective binding sites for the heterochromatin protein 1, which facilitates heterochromatin formation. Finally, the RITSC promotes RNA-pol-II release via an unknown mechanism and represses gene activity. Because the TGS mechanism appears to be evolutionarily conserved, it has been proposed to function in genome maintenance and DNA repair as additional roles to gene regulation [194, 214-216].

1.6.2 RNAi therapeutics for β-haemoglobinopathies

The ability of siRNAs to induce specific and highly efficient silencing of virtually any gene was harnessed for the development of siRNA-based therapeutics towards the treatment of essentially any gene-causing disease [217]. A search for ‘siRNA’ as a keyword online at https://clinicaltrials.gov/ retrieved information on 39 completed, ongoing or pending human clinical trials worldwide. In order to achieve clinical efficacy, siRNAs need to be delivered inside of target cells for RNAi processing. When administered systemically, the efficiency of the siRNA delivery system to reach the target site is limited by physiological barriers impeding siRNA entry into cells. Therefore synthetic siRNA duplexes are engineered to include chemical modifications and/or material conjugates for target-specific delivery [218]. In this context, a recent study

reported that systemic administration of siRNA-conjugated lipid nanoparticles targeting hepatic *Tmprss6* expression in TI mice relieved suppression on the iron-regulatory hormone hepcidin and reduced iron overload which, ultimately, improved erythropoiesis, anaemia and other clinical sequelae [219]. In an alternative approach, bacteria- or virus-mediated RNAi delivery systems are engineered to produce small-hairpin RNAs (shRNAs) which are processed into functional siRNAs intracellularly. The shRNA sequences mimic conserved pre-miRNA features and, as such, are processed by the endogenous miRNA pathway. Transcription produces a hairpin consisting of two complementary 21–29nt siRNA sequences linked by a 6–9nt loop, and featuring a 2nt 3' overhang. The resulting hairpin is exported into the cytoplasm and feeds into the RNAi pathway at the Dicer processing stage (**Figure 1.13**) [205, 220]. LVs are better suited to attain long-term shRNA expression and durable gene silencing because the viral DNA is integrated into the genome of the host cell [205]. Based on their capacity to confer sustainable RNAi, shRNA-expressing constructs have frequently been used for *in vivo* gene silencing. Xie and colleagues demonstrated that shRNA-targeted reduction of excess HBA chains in thalassaemic mice (Hbb^{th4}/Hbb^{+}) carrying a human splicing-deficient *HBB* allele (βIVS-II-654), improved erythropoiesis and sustained amelioration of anaemia [221]. Further studies using HBA-specific siRNAs restored HBA/HBB chain balance in murine thalassaemic primary erythroid cell cultures, demontrating the therapeutic potential of HBA suppression for the treatment of β-thalassaemia [222].

The shRNAs expressed from RNA-pol-III promoters, such as U6 and H1, have been widely used in a variety of mammalian cell types because they provide constitutive high expression levels necessary for potent target gene knockdown. However, RNA-pol-III promoters do not provide the spatial or temporal control required for *in vivo* RNAi therapy. In order to achieve lineage-specific targeting of mRNAs and reduce the likelihood of toxic side-effects and off-target gene expression knockdown associated with upregulated siRNA production and saturation of the cellular RNAi machinery, expression of shRNA via RNA-pol-II promoters is required [205, 223]. To this end, Sadelain and colleagues [224] expressed shRNAs from within the intron of an *HBG* transgene linked to the RNA-pol-II *HBB* promoter as a therapeutic strategy for SCD. The shRNA was expressed from the IVS-II region resulting in successfully knockdown of its target β^{S}

mRNA in progeny of $CD34^+$ cells. Importantly, shRNA expression from a position near the branch site of IVS-II presented higher RNAi efficiency without adverse effects on *HBG* transgene expression in comparison to the other two candidate hairpin positions (**Figure 1.14**). However, in spite of different levels of knockdown, none of these positions resulted in complete loss of functionality. This study provided proof-of-principle for synergistic gene delivery and lariat-encoded RNAi in human $CD34^+$ cells [224]. Of note, another study by Lin and colleagues [225] reported gene silencing with concurrent mRNA production using an artificial splicing-competent intron containing inserts of sense, antisense or hairpin RNAi-encoding sequences embedded in a position between the splice donor site and branch-point domain [225].

The structural criteria for predicting effective knockdown with shRNAs expressed from RNA-pol-II promoters are poorly understood, with only rare references to such constructs found in the literature [223, 226]. As an alternative to improving this system, synthetic shRNA stems are embedded within an miRNA target site context [220, 227]. Essentially, the siRNA sequences of the shRNA stem are incorporated into the miRNA scaffold in place of the natural target sequences, leading to creation of desired siRNA products. The miRNA-embedded shRNAs, termed 'shRNA-miRs', closely resemble endogenous triggers and serve as natural substrates in the miRNA biogenesis pathway [228]. To date, there are just a few reports describing efficient knockdown from RNA-pol-II-expressed shRNAs but which exhibit lower efficiency compared to RNA-pol-III-expressed shRNAs [220, 229-232]. In addition, the design parameters proposed for achieving superior knockdown were neither consistent across these studies nor did they predict consistent shRNA potency when considering different targets. To date, the rules for efficient transcription of shRNAs from RNA-pol-II promoters are still being developed and refined [209, 228]. A recent study demonstrated that RNA-pol-II-expressed BCL11A-specific siRNAs derived from a miR223 context produced lower *BCL11A* knockdown compared to RNA-pol-III-expressed shRNA constructs, possibly owing to variation in the processing sites of flanking sequences. In consequence, the Dicer cleavage site was shifted to produce guide strand sequences varying by 2–4nt and thus conferring different knockdown potency. Engineering the guide strand sequence of the shRNA-miR to accommodate the shift in cleavage sites led to production of RNAi triggers identical to RNA-pol-III-derived shRNAs, which gave rise to a >100-fold reduction

in BCL11A mRNA, albeit not to the level produced by RNA-pol-III-derived shRNAs. Applying the optimized shRNA-miR strategy to human erythroid progenitor cells reduced *BCL11A* expression which led to *HBG* reactivation and formation of substantial amounts of HbF [233]. Therefore it would appear that the optimization of stem sequences represents a possible strategy for the design of shRNA-miR therapeutics.

An alternative approach to therapy involved the design of an artificial ZF transcriptional activator termed GG1-VP64 to interact with the proximal *HBG* promoters and guide gene reactivation. In addition to activating the *HBG* promoters and/or perturbing repressor binding locally at these sites, it was proposed that GG1-VP64 could possibly modulate long-range chromosomal interactions to promote interaction between the *HBG* promoters and βLCR [234]. Treatment of BM-derived $CD34^+$ cells from transfusion-dependent, β^+-thalassaemic patients resulted in elevated HbF production (55%). Combining endogenous activation of *HBG* loci with *HBG* transgene addition further increased HbF production to approximately 60%, a level that in all likelihood would be therapeutic in a β-thalassaemia and SCD context [111, 235].

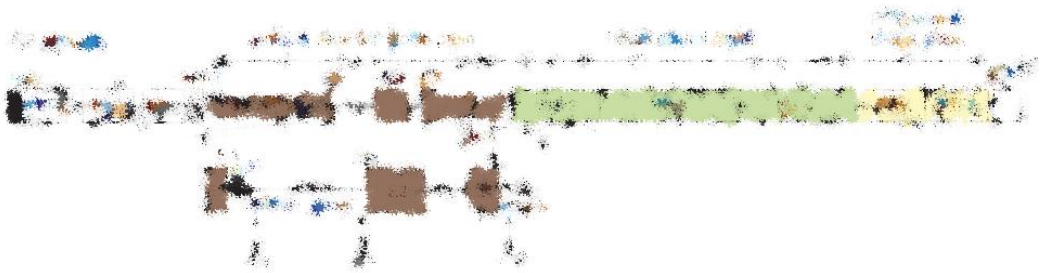


Figure 1.14. Schematic drawing of the G9 lentiviral vector expressing β^S-specific shRNAs from the intronic region of recombinant *HBG*. The shRNA design consists of a simple hairpin structure of a 19bp stem linked by a 9bp loop. The three candidate shRNA cloning sites included i) an undisclosed position immediately upstream the lariat branch site, ii) near the splice donor site at the 5' end of IVS-II, and iii) within the 5' UTR. *HBG* (*HBG2* or *HBG1*) used as the expression platform is not defined. As reasoned from the choice of restriction sites used for cloning, additional sequences must have been artificially included in the intron, although not specified. Reproduced from [224].

1.7 Gene therapy clinical trials for β-thalassaemia

The first human gene therapy (GT) trial for β-thalassaemia (June 7th 2007; France), involving three compound heterozygous β^E/β^0-thalassaemic patients and using an LV (LentiGlobin HPV569) containing a βLCR HS2/HS3/HS4-mini-βT87Q-globin gene flanked by cHS4 insulator elements embedded within the U3 regions of the vector long terminal repeats (LTRs), was reported by Leboulch and colleagues in 2010 [236]. Patients received gene-modified autologous HSCs following full-myeloablative conditioning (Busulfan 14 mg/kg). The first patient received unmanipulated back-up cells early-on due to engraftment failure. The third patient showed low gene marking in circulating blood cells and was thus maintained on transfusion therapy. However, the second patient showed a gradual increase in Hb levels reaching 90–100 g/L and has been transfusion-independent for over two years. Of note is that the pool of haemoglobin within this patient was made up in equal proportions of LV-derived HBB and the endogenous HbF and HbE [236]. The sub-optimal vector performance, which led to low levels of vector-derived protein was associated with low transduction efficiency (about 30% [237]) as well as low engraftment of gene-corrected HSCs (10%–20% [238]). As a result, expression of endogenous Hbs was crucial for attaining therapeutic success. In addition, ~50% of the vector-derived HBB protein was produced from a dominant, myeloid-biased cell clone harbouring an LV insertion within the *HMGA2* proto-oncogene. This clonally dominant myeloid cell clone carried an LV insertion within the third intron of *HMGA2* and resulted in aberrant splicing between the third exon and vector LTR leading to overexpression of a truncated (exons 1-3) *HMGA2* transcript and protein. Although there is evidence that such a truncated HMGA2 is oncogenic, this led to a clonal yet benign cell expansion maintained at ~5% of the circulating nucleated cells [236, 237]). The LentiGlobin HPV569 vector contains two copies of the 250bp core element of the cHS4 insulator in the 3' LTR that are duplicated in the 5'LTR upon provirus integration following transduction.

It was proposed that the cHS4 duplication triggered the recombination, which led to deletion of one of the two elements and subsequent loss of insulating activity, in turn possibly contributed to *HMGA2* upregulation via the βLCR elements in the LV [180]. Overall, this first GT trial was

considered a clinical success because the treated patient no longer needs transfusion therapy and does not show a malignant or pre-malignant state. However, given that these outcomes would not have been achieved in the absence of contributions from HbF and HbE as well as that from the clonally expanded myeloid precursor, this case can be seen at best as a partial success. It is clear that many improvements to procedure and vector are still required to ensure efficacy and safety in future trials.

A second clinical trial targeting β-thalassaemia was launched in July 2012 in the USA by Sadelain and colleagues. As yet, no study findings have been published. The strategy was briefly described in [239] and entailed PB $CD34^+$ HSCs mobilization by granulocyte colony-stimulating factor (G-CSF) for *ex vivo* treatment with the TNS9.3.55 vector. Essentially, the TNS9.3.55 vector is an improved version of the previously described TNS9 LV [183, 184], which received small unpublished modifications to increase efficiency of production and thus the titre without, altering the therapeutic transcription unit [239]. A recent publication by the same group [240] disclosed the findings of a pilot study, which was performed using *in vitro* model systems to assess protocol efficiency and safety. The results confirmed successful mobilization of $CD34^+$ cells in four patients with TM in amounts sufficient for transduction and reinfusion ($8x10^6$ $CD34^+$ cells/kg). Using clinical grade TNS9.3.55 vector, $CD34^+$ cells were transduced under optimized culture conditions in three validation runs with an average VCN of 0.53, and vector-encoded HBB protein was measured by HPLC at 73%–100% of normal hemizygous HBB output**.** The results were confirmed *in vivo* by a xenograft approach in NSG mice [240]. The outcome suggested a complete correction of β^+-thalassaemia and a promising therapeutic effect for β^0-thalassaemia [240]. Recent reports documented that the three β^0/β^+ thalassaemic patients treated with gene-augmented autologous HSCs after reduced chemical conditioning (Busulfan 8 mg/kg) remain transfusion-dependent after one year of treatment but with increasing intervals between transfusions [241].

A third clinical trial using an improved LentiGlobin vector design (LentiGlobin BB305, [242]) is currently in progress for β-thalassaemia and SCD in France and the USA. This trial is sponsored by Blue Bird BIO. Modifications in vector design included replacement of the 5’ U3

LTR promoter/enhancer with the CMV promoter/enhancer and removal of the cHS4 insulator elements, without altering the therapeutic transcription unit, in an effort to improve viral titres, transduction efficiency and safety. Preliminary clinical results have been announced for two patients with TM with the β^E/β^0 genotype, who were transfusion-independent at 4.5 and 2 months post GT with Hb levels at 10.1 g/dL and 11.6 g/dL of which 6.6 g/dL (VCN 1.5) and 4.7 g/dL (VCN 2.1) was vector-derived HBB. Early results are available on-line at: http://investor.bluebirdbio.com/phoenix.zhtml?c=251820&p=irol-newsArticle&ID=1939867.

Additional clinical trials targeting both β-thalassaemia and SCD at other centres are planned (e.g., see https://clinicaltrials.gov/ct2/show/study/NCT02186418).

1.8 Aims and Objectives

Gene therapy by *ex vivo* genetic modification of haematopoietic stem and progenitor cells (HSPCs) represents a promising therapeutic option for β-thalassaemia, a severe hereditary anaemia. As described above (**sections 1.6 and 1.7**) LVs have been used extensively to transfer functional copies of *HBB* to HSPCs, the endogenous counterpart of which is defective. The aim of gene augmentation therapy is to increase the amount of normal gene product to a level where the normal phenotype is restored. Importantly, in β-thalassaemia, for a curative therapeutic effect, a minimum of 50% of normal levels of HBB is required, which is currently unattainable by extant vector systems. In this study, we aim to develop an advanced vector system with ability to provide cure for β-thalassaemia major with greater vector-mediated therapeutic efficiency than current LV designs. Our work is based on a well-characterized, efficient, and safe HBB-expressing vector called 'GLOBE' [156, 188], to develop a range of LVs encoding shRNAs against the *HBG* suppressor BCL11A. Using this combinatorial expression system of vector-derived and endogenous globins we aim for a therapeutic effect per LV integration event far greater than that of any other current system. To test the efficiency of the novel vector design more rapidly, green fluorescent protein (GFP)-specific shRNAs will be designed to target this reporter gene and provide preliminary insight for synergistic gene delivery and lariat-encoded RNAi. A GFP-expressing cell line with long-term *GFP* expression will be produced as a research tool for the analysis of shRNA-mediated GFP knockdown.

Essentially, all LVs will be assessed for safety and efficacy in relevant *in vitro* cell lines for which optimal erythroid differentiation conditions will be established. The BCL11A-targeting LVs will be tested in K562 and HEL erythroid cell lines, which display an embryonic and/or fetal pattern of HBB-like chain production. The key goal of this study is to functionally characterize these vectors in normal primary human erythroid cell progenitors, followed by therapeutic assessment of selected vectors in β-thalassaemia patient samples. The performance of these vectors will be compared to the original GLOBE vector. Extraction protocols of HSPCs from PB samples will be established and based on density-gradient centrifugation and simple ACK lysis of red blood cells. Lastly, a clinically relevant primary human cell culture model system will be established that can support *in vitro* expansion and erythroid differentiation of naïve and transduced $CD34^+$ cells.

2.0 Materials and Methods

2.1 DNA extraction

2.1.1 Genomic DNA (gDNA) extraction from eukaryotic cells

2.1.1.1 DNA extraction using simple cell lysis

The procedure was adapted from [243] and was used to extract gDNA from transduced HEL cells (**section 2.13.3**) for biological titration using qPCR. Cells (5–10x10^6) were suspended in 500 µL lysis buffer supplemented with 100 µg/mL proteinase K (*Sigma-Aldrich, St. Louis, USA*), followed by incubation (3 hours–overnight) at 56 °C with occasional vortexing. DNA was precipitated with 0.7 volumes of room temperature (RTemp) isopropanol, followed by centrifugation at 16,100 g (maximum speed) for 20 minutes. The precipitated DNA was washed with 1 mL pre-chilled 75% ethanol and recovered by centrifugation at 16,100 g (maximum speed) for 10 minutes. The pellet was re-suspended in 200–500 µL 10 mM Tris.HCl, pH 8.5. Viscous DNA was fragmented by exposure to brief sonication for 5 seconds at 5 µm amplitude using a microtip probe (VirTis VirSonic, *Boston Laboratory Equipment, MA, USA*).

Lysis buffer	100 mM Tris.HCL pH 8
	5 mM EDTA
	100 mM NaCl
	0.2% SDS

2.1.1.2 DNA extraction using phenol:chloroform:isoamyl alcohol

Transduced $CD34^+$ cells (0.5–2x10^6) were suspended in 400 µL solution A supplemented with 20 µL 10% SDS and 20 mg/µL proteinase K, vortex-mixed and incubated at 65 °C for 2 hours. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1 by volume, *USB, OH, USA*) was added to the lysed cells and mixed thoroughly prior to centrifugation at 16,100 g (maximum speed) for 3 minutes. The upper aqueous DNA-containing phase was removed to a fresh tube, mixed with 2.5 volumes of 100% chloroform and centrifuged at 16,100 g (maximum speed) for 5 minutes to remove residual phenol. The upper phase, containing purified nucleic-acid, was

transferred to a fresh tube and mixed with 0.7 volumes of isopropanol, followed by a pre-chilled 75% ethanol wash, both at 16,100 g (maximum speed) for 5 minutes.The pellet was re-suspended in 30 µL DNase-free water.

Solution A	10 mM Tris.HCl, pH 8
	2 mM EDTA
	400 mM NaCl

2.1.1.3 DNA extraction using FlexiGene DNA kit

DNA from transduced HEL and K562 cells was extracted using the FlexiGene DNA kit (*Qiagen, Hilden, Germany*) according to manufacturer's instructions. Briefly, 1–2x10^6 cells were suspended in lysis buffer FG1 followed by centrifugation at 10,000 g for 20 seconds. The pellet, mainly of nuclear and mitochondrial content, was re-suspended in a denaturing guanidine-hydrochloride-containing buffer supplemented with proteinase K and incubated at 65 ºC for 10 minutes. DNA was precipitated with 0.7 volumes of isopropanol and recovered by centrifugation. The precipitated DNA was washed with 1 mL pre-chilled 75% ethanol and collected at 16,100 g (maximum speed) for 10 minutes following re-suspension in 150 µL hydration buffer FG3 (10 mM Tris.HCl, pH 8.5).

2.1.2 Plasmid DNA extraction from bacterial cells

2.1.2.1 Purification of plasmid DNA using phenol:chloroform

Plasmid DNA from bacterial mini cultures (**section 2.12.4**) was extracted using alkaline lysis followed by phenol:chloroform purification. The bacterial cell pellet was re-suspended in 300 µL of solution 1 supplemented with 100 mg/mL RNase (*Sigma-Aldrich*) and lysed with 300 µL of solution 2, followed by gentle inversion and incubation at RTemp for 5 minutes. Addition of 300 µL of pre-chilled neutralization solution 3 and incubation of the mixture on ice for 15 minutes, caused SDS to precipitate pulling down proteins, gDNA and cell debris upon centrifugation at 16,100 g (maximum speed) for 20 minutes at 4 °C. Phenol:chloroform:isoamyl alcohol (*USB*)

extraction was performed as before (**section 2.1.1.2**), but instead using an equal amount of cholorofom to remove residual phenol. Plasmid DNA was re-suspended in 50–100 µL of 10mM Tris.HCl, pH 8.5.

Solution 1	0.06 g/L Tris 3.72 g/L $Na_2EDTA.2H_2O$, pH 8
Solution 2	200 mM NaOH 10% SDS
Solution 3	294.5 g/L potassium acetate pH 5.5 with glacial acid

2.1.2.2 Silica-based plasmid DNA purification

Plasmid DNA from bacterial maxi cultures (**section 2.12.4**) was extracted using the Macherey-Nagel NucleoBond maxi kit (*Macherey-Nagel GmbH, Duren, Germany*) according to manufacturer's instructions. The kit uses an alkaline lysis method adapted from [244], followed by binding to and elution of plasmid from an anion-exchange chromatography column. Briefly, the bacterial cell pellet was re-suspended in RNase A-supplemented buffer and lysed by NaOH/SDS treatment. The lysate was neutralized in potassium acetate buffer causing DNA to precipitate and bind to the positively-charged silica membrane. Plasmid DNA was eluted under alkaline conditions and washed with 0.7 volumes isopropanol. Precipitated plasmid DNA was washed in 2.5 volumes of 100% ethanol supplemented with 0.1 volumes of 3 M sodium acetate, followed by an additional wash with 1 mL pre-chilled 75% ethanol, both at 16,100 g (maximum speed) for 10 minutes. The plasmid DNA was re-suspended in 300–500 µL of 10mM Tris.HCl, pH 8.5.

2.1.3 DNA quantitation and purity assessment

DNA concentration was quantitated using a NanoDrop 1000 spectrophotometer with ND-1000 v3.8.1 software (*Thermo Scientific Inc., MA, USA*). DNA purity was determined by optical density readings (OD) at 260, 280 and 230 nm wavelength. An OD260/OD280 ratio ≥1.8 and a secondary OD260/OD230 ratio ≥2 indicated a pure DNA sample.

2.2 Agarose gel electrophoresis of DNA

Agarose gels were prepared at concentrations of 1%–2%, depending on the size of the bands to be separated by dissolving the appropriate amount of agarose powder (*Sigma-Aldrich*) in 100 mL of 1X Tris-Borate-EDTA (TBE) buffer. DNA samples were mixed with 10% of 10X bromophenol blue (BPB) / xylene cyanol (XC) loading dye and were loaded on the gel, along with 0.5 µg of 100bp or 1kb DNA ladder (*NEB®; MA, USA*) (**Figure 2.1**). For reference, BPB and XC run at approximately equal rates to 300 and 4000 bp DNA fragments, respectively, in 1% agarose. Gels were run at 90–115 V/cm until BPB had migrated ¾ of the gel length. Gels were stained in the intercalating agent ethidium bromide (0.5 µg/mL), DNA fragments visualized with a UV transilluminator (*Vilber Lourmat, Wembley, Australia*) and images captured using the Mega-Capt software.

10X TBE	108 g Tris 55 g boric acid 40 mL 0.5 M EDTA pH8.0
10X BPB / XC loading dye	2.5 g Ficoll Type 400 25 mg bromophenol blue 25 mg xylene cyanol FF

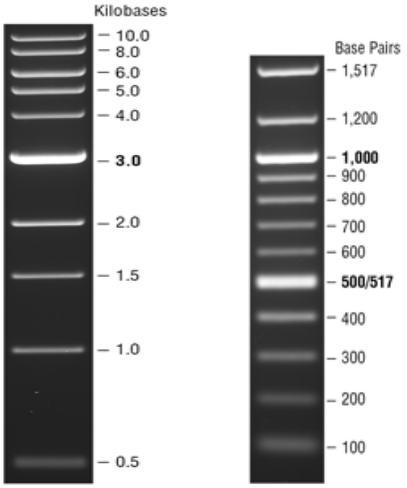


Figure 2.1. Images of 1 kb DNA ladder (left) and 100 bp DNA ladder (right) retrieved from the *NEB*® official site.

2.3 DNA sequencing

The sequencing reactions were performed using purified plasmid DNA as template and the BigDye Terminator v1.1 Cycler sequencing kit (*Applied Biosystems, MA, USA*). The sequencing cocktail consisted of 4 µL BigDye v1.1, 4 µL 5X BigDye buffer, 4 µL 5X GC-RICH solution (*Roche, Basel, Switzerland*), 250 nM primer, and 800 ng template in a final reaction volume of 20 µL. Cycle sequencing was performed on a Tgradient Thermocycler (*Biometra GmbH, Goettingen, Germany*) using the following conditions: 1 minute 96 °C, (10 seconds 96 °C, 5 seconds 50 °C, 4 minutes 60 °C) x 25 cycles, 15 °C soak. Purification of extension products was carried out with Performa® DTR Gel Filtration Cartridges (*Edge Biosystems, Maryland, USA*) as described by the manufacturer. Sample electrophoresis and data analysis were performed on a Hitachi 3031xl Genetic Analyzer with Sequence Detection Software version 5.2 (*Applied Biosystems*). Sequencing primers were as follows:

Name	Sequence 5' to 3'
DGGE Fragment 1 R	CAACTTCATCCACGTTCACC
ARMS 10	ATGCACTGACCTCCCACATTCCCTTTTTAG
αPCO4_FW	GGTGAACGTGGATGAAGTT
MA821_T87Q_SEQ_FW	GGACAGCTAGGTGGCAAAAG
MA821_T87Q_SEQ_RV	AGAAAACATCAAGGGTCCCA
MA821_NotI_FW	ACCCTCTATTGTGTGCATCAAAGG
MA821_ClaI_RV	GAAGAGGAAGGCATGAAAACAGG
3402_SEQ	AGAACAATTTGCTGAGGGCT
3099_SEQ	GGAAGAGCAAAACAAAAGTA
3861_SEQ	TGAATAGAGTTAGGCAGGGA
IVS2_shSEQ_FW	CTCGGTGCCTTTAGTGATGG
IVS2_shSEQ_RV	AGTGATACTTGTGGGCCAGG
pLKO.1_SEQ	CAAGGCTGTTAGAGAGATAATTGGA

2.4 RNA extraction

2.4.1 RNA extraction using PeqGold Trifast™ solution

RNA was extracted from experimental samples of haematopoietic cell lines, K562 and HEL, using PeqGold TriFast™ (*PeqLab, Erlangen, Germany*) according to manufacturer's instructions. Briefly, ≥3x10^6 cells were suspended in 1 mL TriFast™ solution and incubated at RTemp for 5 minutes prior to phase separation. Alternatively, TriFast™-lysed cells were stored at -80 ºC and thawed on ice before processing. Samples were mixed with 0.2 mL chloroform, incubated at RTempfor 5 minutes and centrifuged at 12,000 g for 5 minutes. The mixture separated into the upper aqueous phase, the interphase and the lower phenol-chloroform phase containing the RNA, DNA and proteins, respectively. RNA in the aqueous phase was precipitated in 0.5 mL isopropanol at 12,000 g for 10 minutes at 4 ºC, and was washed twice

with 75% chilled ethanol at 12,000 g for 10 minutes at 4 ºC. The RNA pellet was air-dried and re-suspended in RNAse-free water.

2.4.2 Silica-based RNA purification

RNA was extracted from primary human $CD34^+$ cells using the RNeasy mini kit (*Qiagen*), which includes binding to and elution of RNA from silica-based membrane columns, in accordance with manufacturer's instructions. Briefly, ≥$0.5x10^6$ cells were lysed in a denaturing guanidine-thiocyanate-containing buffer supplemented with β-mercaptoethanol (*Sigma-Aldrich*) to facilitate inactivation of RNases. Lysates were mixed with an equal volume of 70% ethanol and were washed through the silica-based membrane at 16,100 g (maximum speed) for 15 seconds. The RNA was selectively bound to the membrane while DNA and other contaminants were removed with subsequent washes. RNA was eluted in RNase-free water.

2.4.3 RNA quality assessment

An aliquot of the RNA sample (≥200 ng) was analysed using 1% agarose gels (**section 2.2**). RNA integrity was confirmed by the production of two bands corresponding to the 28S and 18S rRNAs at an approximate intensity ratio of 2:1, whereas degraded RNA presented with a smeared appearance and correspondingly lower ratio. Assessment for DNA contamination in RNA preparations was based on reverse-transcriptase qPCR (RT-qPCR) of minus-RT control reactions for each RNA sample (**section 2.6.1**). Signal detection in PCR using minus-RT control samples that were not reverse transcribed would suggest the presence of gDNA serving as a template for product amplification.

The purity of RNA was assessed by spectrophotometry (**section 2.1.3**)

2.5 Protein analysis

2.5.1 Western blot analysis

2.5.1.1 Sample preparation and SDS-PAGE electrophoresis

Cells were washed in ice-cold PBS at 300 g for 5 minutes at 4 °C and suspended in RIPA buffer supplemented with protease inhibitors (*Roche, Basel, Switzerland*). Cell lysates were incubated on ice for 15 minutes and cleared by centrifugation at 15,700 g for 15 minutes at 4 °C. The protein-containing supernatant was mixed with equal volumes of 2X loading buffer and stored at -20 °C. Samples were thawed on ice and heated at 95–100 °C for 5 minutes before use. The protein extract was separated on SDS-PAGE gels consisting of the lower resolving gel and the top stacking gel:

Reagents	**Resolving gel**		**Stacking gel**
	8%	**12%**	
dH2O	3.85 mL	4.3 mL	3 mL
40% acrylamide	2.0 mL	3 mL	0.625 mL
1.5 M Tris pH 8.8	2.0 mL	2.5 mL	-
0.5 M Tris pH 6.8	-	-	1.260 mL
20% SDS	80 µL	100 µL	25 µL
30% APS	53.6 µL	67 µL	16.6 µL
TEMED	16 µL	20 µL	5 µL

Volumes of protein extract equivalent to $2x10^5$ lysed cells were loaded, unless otherwise stated, in parallel to a molecular weight protein marker (*Nippon Genetics, Duren, Germany*) for protein size determination. The gels were submerged in 1X running buffer and run at 45 mA until migration of the blue dye front reached the end of the glass plates.

2.5.1.2 Wet Tank blotting

Proteins were transferred from gels to blotting membranes using wet electrophoretic transfer. Briefly, the protein-containing gel and the Nitrocellulose Parablot NCP membrane (*Macherey Nagel*) were sandwiched between two stacks of buffer-wetted filter paper, rolled over to smooth out air bubbles and enclosed in a gel holder cassette. The assembly was placed in a BioRad tank containing plate electrode cards such that the membrane was closest to the positive electrode. The tank was filled with 1X blotting buffer, and placed in an ice block to prevent temperature fluctuations. A voltage of 35–40 V at a constant mA was applied for 2 hours. Protein transfer was assessed by staining membranes with Ponceau Red solution (*Sigma-Aldrich*), producing pale pink bands, and destaining by repeated washing in dH_2O or TBST.

2.5.1.3 Immunoblotting and detection

To block non-specific background-antibody binding, membranes were agitated in 1% BSA (*Sigma-Aldrich*) at RTemp temperature for 1 hour. Blots were removed from blocking buffer, overlaid with 5 mL primary antibody diluted 1:1000 in 1% BSA in a 50-mL tube and incubated overnight (up to 18 hours) at 4 °C on a roller mixer. Residual primary antibody was removed by washing membranes, thrice, with 1X TBST buffer for 10 minutes at RTemp with shaking. Membranes were incubated in fluorescently-labelled secondary antibody diluted 1:1000 in dry milk dissolved in 1X TBST buffer, for 1 hour at RTemp with shaking. Membranes were washed, thrice, with 1X TBST buffer for 10 minutes at RTemp with shaking. Wash buffer was removed and membranes were covered with chemiluminescence staining buffer (Lumisensor, *GenScript, New Jersey, USA*) for 3 minutes in dark. The chemiluminescence signal was detected using UVP Biospectrum 810 Imaging system (*UVP, Cambridge, UK*), and transformed into a digital image for analysis using VisionWorks™ LS v7.1 software.

2.5.1.4 Quantitative densitometry for proteins

Immunoblot digital images were imported into Photoshop CS6 software in order to produce a gray-scale image required for densitometry analysis. The image was saved in TIF format and

opened with the Fiji ImageJ image analysis software. ImageJ converts pixel intensities into optical density. The relative density value of the sample protein and loading-control bands were measured and corrected for background signal, before control bands were used for loading correction. The calculated relative density values are arbitrary numbers not representing specific units of protein. For comparability between experiments (where overall staining intensity may vary depending on incubation time etc.), density values for each membrane were normalised to the mock-treated control sample as 100%.

2.5.1.5 Buffers and solutions

Buffer name	Composition
1.5 M Tris pH 8.8	18.15 g Tris base 100 mL dH2O pH 8.8 with 6 N HCl
0.5 M Tris pH 6.8	6 g Tris base 100 mL dH2O pH 6.8 with 6 N HCl
RIPA buffer	150 mM NaCl 20 mM Tris-HCl, pH 7.4 1% NP40 0.1% SDS 0.5% sodium deoxycholate 5 mM EDTA
2X Sample Buffer	5 mL 0.25 M Tris pH 6.8 2 mL 10 % SDS 2.1 mL Glycerol 0.4 mL β-mercaptoethanol 0.5 mL 0.1 % bromophenol blue

10X Running Buffer	30.2 g/L Tris 144.2 g/L Glycine 1% SDS
10X Blotting Buffer	30.2 g/L Tris 144.2 g/L Glycine
1X Blotting Buffer	100 mL 10X Blotting Buffer 150 mL methanol 750 mL dH2O
10X TBS	87.7 g/L NaCl 12.1 g/L Tris, pH 7.4
1X TBS – 0.05% Tween 20 (TBST)	100 mL 10X TBS 5 mL 10 % Tween 20 895 mL dH2O

2.5.1.6 Antibodies

haemoglobin β (37-8) mouse monoclonal IgG1; sc-21757 haemoglobin γ (51-7) mouse monoclonal IgG1; sc-21756 actin (C-2) mouse monoclonal IgG1; sc-8432 GAPDH (6C5) mouse monoclonal IgG1; sc-32233 Bcl-11a (H-116) rabbit polyclonal IgG; sc-366825	*Santa Cruz Biotechnology, CA, USA*
Bcl11a rabbit polyclonal rabbit polyclonal IgG; A300-382A	*Bethyl Laboratories, TX, USA*
Anti-Ctip1 antibody [15E3AC11]; ab18688 (anti-BCL11A) Anti-GFP rabbit polyclonal; ab6556	*Abcam, Cambridge, UK*
peroxidase-conjugated AffiniPure goat anti-mouse IgG (H=L); 115-035-003 peroxidase-conjugated AffiniPure goat anti-rabbit IgG (H=L); 111-035-003	*Jackson ImmunoResearch Laboratories, PA, USA*

2.5.2 Separation of globin chains by reversed-phase liquid chromatography

2.5.2.1 Sample preparation

Samples of 4 µL of peripheral blood (PB) collected in EDTA anticoagulant (*USB*) were diluted in 996 µL dH_2O, vortex-mixed and incubated at RTemp for 10 minutes. For scarce samples, material was collected at 2,000 g for 10 minutes, lyophilized in 100 µL dH_2O and stored at -80 °C followed by thawing on ice. The lysate was collected at 16,100 g (maximum speed) for 10 minutes at RTemp, and 100 µL haemolysate was applied onto the column for analysis.

Samples of erythroid culture cells, enumerated by benzidine assay (**section 2.10.1**), were washed twice in 500 µL PBS at 200 g for 5 minutes at RTemp and were suspended in dH_2O at a concentration of $0.5x10^6$ cells per 50 µL dH_2O. Lysed cells were incubated for ≥4 hours on ice or overnight at 4 °C, followed by centrifugation at 16,100 g (maximum speed) for 10 minutes at 4 °C to remove cellular debris and separate the haemoglobin solution. The supernatant-containing protein extract was stored at -80 °C and thawed on ice before use. The injection volume was equivalent to $0.5x10^6$ lysed erythroid cells.

2.5.2.2 Instrument

Globin chain separation was performed using a Shimadzu LC-20AD chromatographic system (*Shimadzu*, *Kyoto, Japan*), equipped with a CBM-20Alite system controller, a LC-20AD dual-plunger solvent delivery system, a SPD-M20A UV-Vis variable wavelength (190–800 nm range) detector, a SIL-20AC auto-injector, a CTO-20A column oven, and an LC-Solution work station interfaced with the Ver. 1.1 software package.

2.5.2.3 Chromatographic separation

Reversed-phase high-performance liquid chromatography (RP-HPLC) separates peptides on the basis of their hydrophobicity by applying an increasing linear gradient of organic solvent to the chromatographic column, **Figure 2.2**.

The globin chain composition of human $CD34^+$ cells and whole-bood samples was assessed on a Vydac 214TP C4 column (4.6 mm x 250 mm, 10 µm, 300 Å, *GraceVydac, USA*) using a 47-minute linear gradient of acetonitrile-methanol [245]. Solvent A was made from 0.1% trifluoroacetic acid (TFA, *Sigma-Aldrich*) in HPLC grade water (*Merck, Darmstadt, Germany*). Solvent B consisted of acetonitrile (*Sigma-Aldrich*) and methanol (*Merck*) in the proportion of 2:1 (v/v). Elution was obtained at 28 °C with a flow rate of 0.8 mL/minute by increasing solvent B from 52% at 0 minutes to 62 and 100% at 40 and 45 minutes, respectively, and then decreasing it to 52% at 47 minutes. Elution was recorded at 280 nm, with the most hydrophilic peptides eluted first. Alternatively, to enhance baseline resolution and reduce analytical run time without compromising peak symmetry, haemoglobin fractions were separated on a Jupiter C18 HPLC column (4.6 mm x 250 mm, 5 µm, 300 Å, *Phenomenex, Torrance, CA, USA*) using a 14-minute gradient elution, as previously described by [246] with slight adjustment. Solvent B composition was as prepared for reversed-phase C4 column, while Solvent A consisted of 0.1% TFA and 4 mM NaOH in HPLC grade water. The gradient slope program started with 56% solvent B for 2 minutes increasing to 62 and 74% at 8 and 20 minutes, respectively, and then decreasing to 56% at 21 minutes. A flow rate of 1 mL/minute was applied and elution was obtained at 22 °C with quantification at 220 nm.

The percent of HBB, HBG1 and HBG2 chains in the sample was quantified by determining the peak area of each elution peak as a fraction of the peak area of the HBA chain.

$$\underset{\text{(expression ratio)}}{\%\ \text{globin}} = \frac{\text{HBB-like chains}^{(\text{peak area})}}{\text{HBA chains}^{(\text{peak area})}} \times 100$$

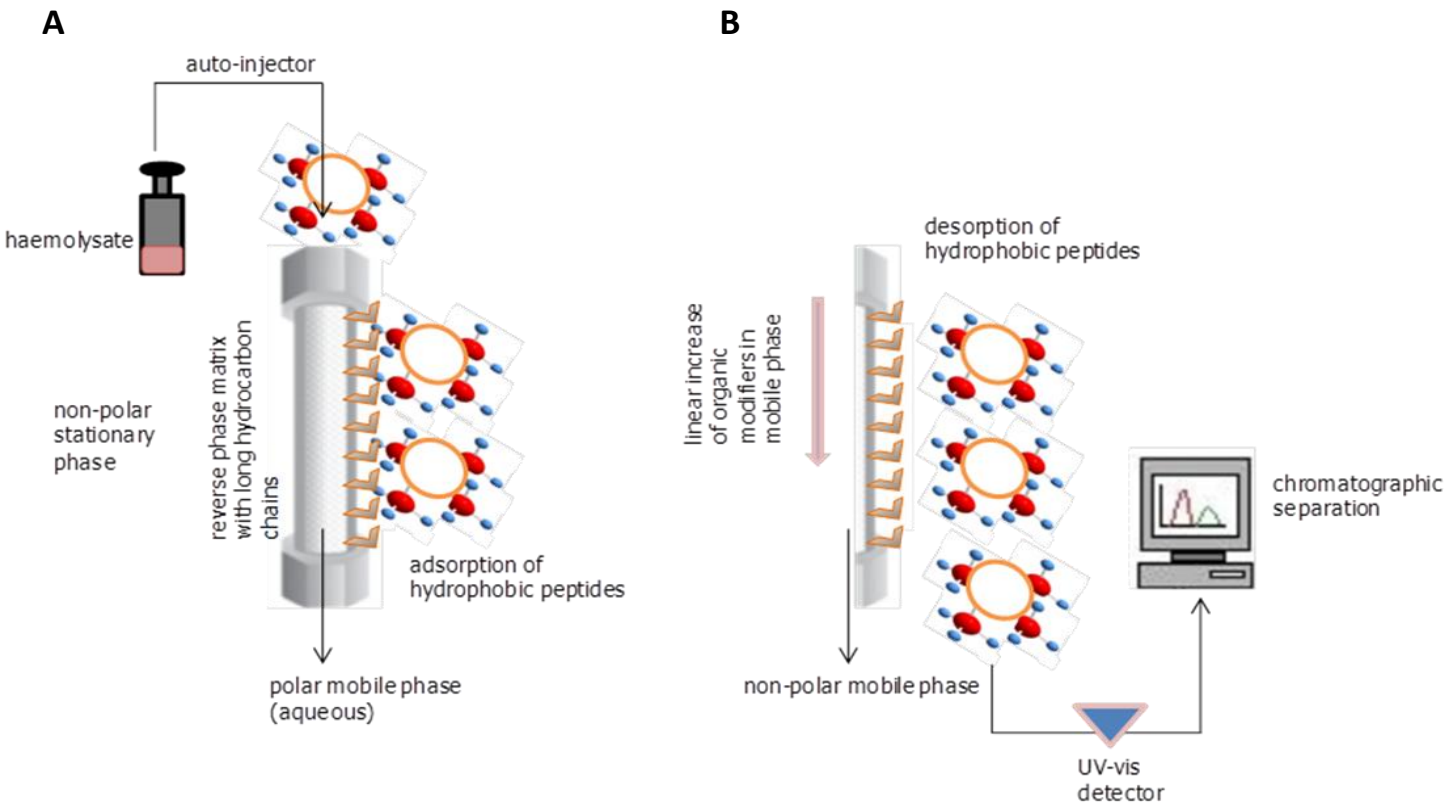


Figure 2.2. Separation of globin chains from biological samples using reversed-phase liquid chromatography. A) Proteins injected into the chromatographic column adsorb to the hydrophobic reverse phase silica matrix through their own hydrophobic moieties under aqueous conditions. **B)** Bound proteins are desorbed from the matrix by decreasing the polarity of the mobile phase through a linear increase in the concentration of organic modifier added to it. Proteins are eluted in order of increasing molecular hydrophobicity and are detected as they emerge from the column. The absorbance signal of each eluate is processed and registered as a peak on the chromatogram of the sample [247, 248].

2.6 Gene expression analysis

2.6.1 Reverse-transcription quantitative PCR (RT-qPCR)

Reverse transcription was performed with the TaqMan® Reverse Transcription Reagents kit (*Invitrogen™, Thermo Schientific*), using total RNA extracted from cells as described in **section 2.4** (**Figure 2.3**). First-strand complementary DNA (cDNA) was synthesized from 200 ng mRNA in a final reaction volume of 10 µL consisting of 2.5 µM random hexamers, 0.4 U/µL RNase inhibitor, 500 µM deoxyribonucleotides (dNTPs), 5.5 mM $MgCl_2$, 1X TaqMan buffer and 1.25 U/µL MultiScribe™ Reverse Transcriptase. The reactions were performed on the Veriti® thermal cycler (*Invitrogen™*), starting with an incubation step at 25 °C for 10 minutes for optimal primer binding, followed by reverse transcription at 48 °C for 30 minutes and enzyme inactivation at 95 °C for 5 minutes. Parallel reactions for each RNA sample were run in the absence of MultiScribe™ enzyme (minus-RT control) to assess gDNA contamination in RNA preparations.

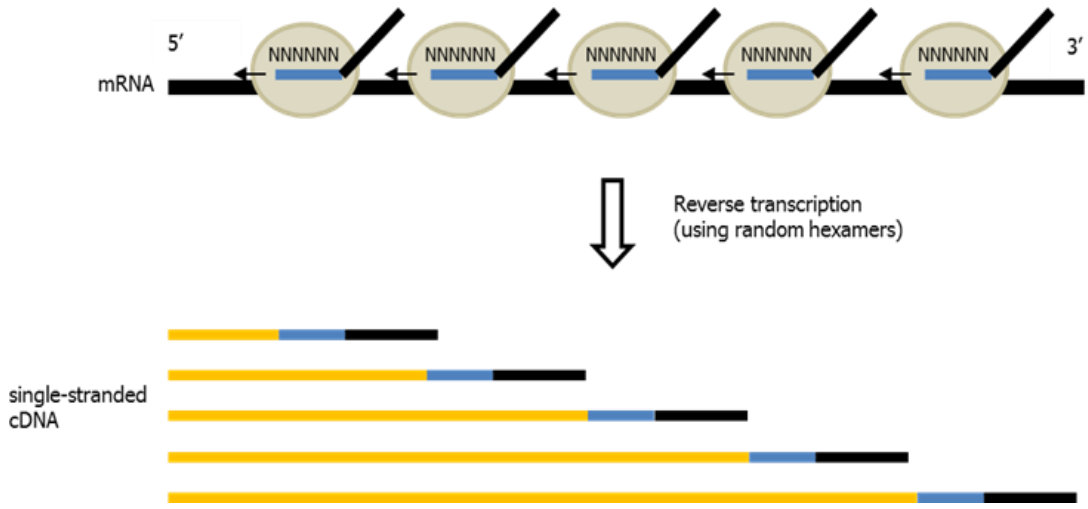


Figure 2.3. Schematic of reverse-transcription qPCR using random hexamers. RNA is reverse transcribed to cDNA using random-hexamer primers (NNNNNN), enabling arbitrary binding to any RNA sequence, with bias towards the 5' end of RNA sequences, but unlike oligo(dT) primers without bias for dA-rich sequence stretches and allowing the use also of rRNA sequences as calibrators.

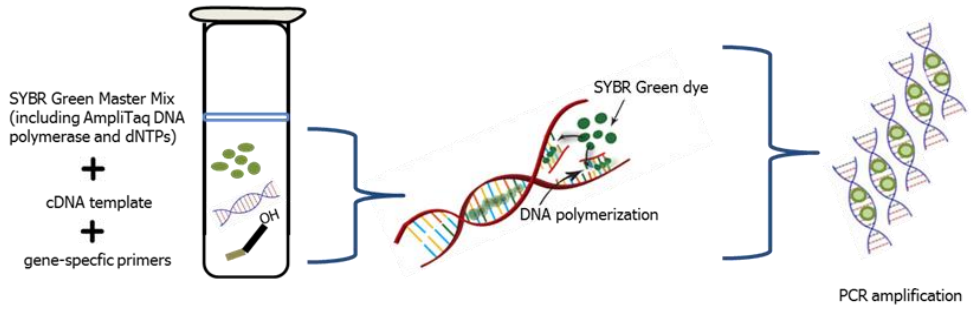
2.6.2 qPCR amplification

Gene expression quantitation was performed on a 7900HT Fast Real Time PCR System (Applied Biosystems) using SYBR Green I chemistry containing a ROX passive reference dye (**Figure 2.4A**–**4B**). The PCR reaction consisted of 12.5 µL of SYBR Green PCR Master Mix (Applied Biosystems), 300 nM of forward and reverse primers (*Metabion International AG*, *Planegg, Germany*), and 2 µL of 1:10-diluted template cDNA in a total volume of 25 µL. The thermal cycling profile started at 95 °C incubation for 10 minutes, in order to activate DNA polymerase, followed by 40 cycles of denaturation at 95 °C for 15 seconds, and combined annealing/extension at 60 °C for 1 minute for the GAPDH, HBG, HBB, GCB1, BCL11A and GFP primer sets, and at 68 °C for 15 seconds for the T87Qx2x3 primer set. Fluorescence was read during the annealing/extension step, with exception of the HBB, GCB1, and T87Qx2x3 assays, the PCR cycling profiles of which incorporated a third segment with fluorescence acquisition at 80 °C for 1 minute (for HBB and GCB1) and at 84 °C for 15 seconds (for T87Qx2x3). The fluorescence acquisition in the additional segment aims to eliminate the non-specific fluorescence signal derived from primer-dimers during the quantification procedure [249]. Melting analysis was performed immediately after PCR to assess for primer-dimer artifacts by heating the reaction mixture to 95 °C for 15 seconds, cooling to 60 °C for 15 seconds and then slowly heating again to 95 °C at a ramp rate of 2% maximum speed for 15 seconds with continuous fluorescence detection (**Figure 2.4C**). Each sample was run in duplicate with non-template controls (NTCs) included in each reaction. Data were analyzed by 7900 Fast System SDS 2.4 software (Applied Biosystems), with automatically set baseline and fluorescence threshold for C_T determination. Gene-specific reverse transcription primers are shown in **Table 2.1**.

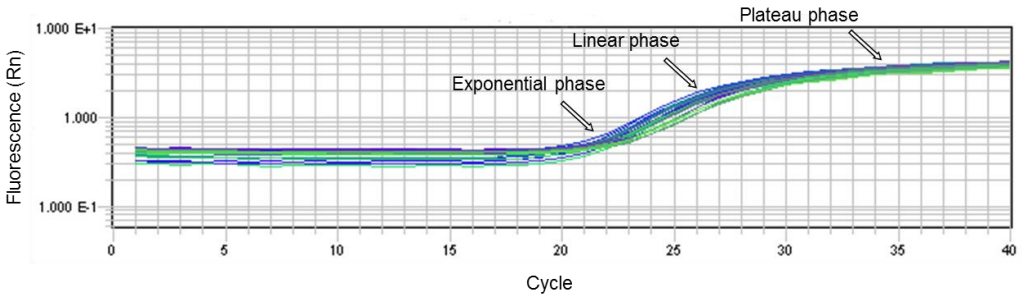
Gene		Oligo	Sequence 5' to 3'	Tm(°C)
GAPDH		GAPDH_FW	TGCACCACCAACTGCTTAGC	53.7
		GAPDH_RV	GATGAGTGCAGAATATGCCCCG	58.1
HBG		HBG_FW	TGACAAGCTGCATGTGGATC	51.6
		HBG_RV	TTCTTTGCCGAAATGGATTGC	56.5
BCL11A	shB9	shB449_FW	GATGAGTGCAGAATATGCCCCG	58.1
		shB449_RV	TGTTCTGTGCGTGTTGCAAGAGA	58.2
	shB1	shB451_FW	CGAGCACAAACGGAAACAATGC	55.6
		shB451_RV	TCCTGTTTGGGGCAAATTCCTC	58.8
HBB		GCB1_FW	TTCACTAGCAACCTCAAACAGACACC	57.6
		GCB1_RV	GATCCCCAAAGGACTCAAAGAACC	58.0
Transgenic HBB		T87Qx2x3_FW	GTGATGGGCCAGCACACAGA	56.5
		T87Qx2x3 _RV	AAGGGCACCTTTGCCCAG	54.8
Endogenous HBB		HBB_FW	GGGCACCTTTGCCACAC	51.5
		HBB_RV	GGTGAATTCTTTGCCAAAGTGAT	54.0
GFP		GFP_FW	GAATCACCGACCTCTCTCCCCA	58.9
		GFP_RV	ACTTCAGGGTCAGCTTGCCGTAG	58.5

Table 2.1. Primer sequences for mRNA quantification. FW: forward; RV: reverse; GAPDH: glyceraldehyde 3-phosphate dehydrogenase; HBB: sumtotal of endogenous and transgenic (HBB^{T87Q}).

A


SYBR Green Master Mix
(including AmpliTaq DNA
polymerase and dNTPs)
+
cDNA template
+
gene-specfic primers
OH
SYBR Green dye
DNA polymerization
PCR amplification

B


Plateau phase
Linear phase
Exponential phase
1.000 E+1
1.000
1.000 E-1
Fluorescence (Rn)
0
5
10
15
20
25
30
35
40
Cycle

C

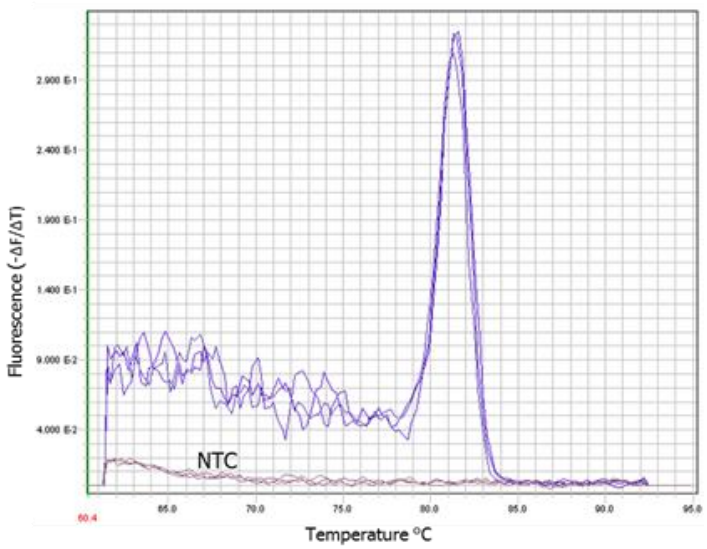
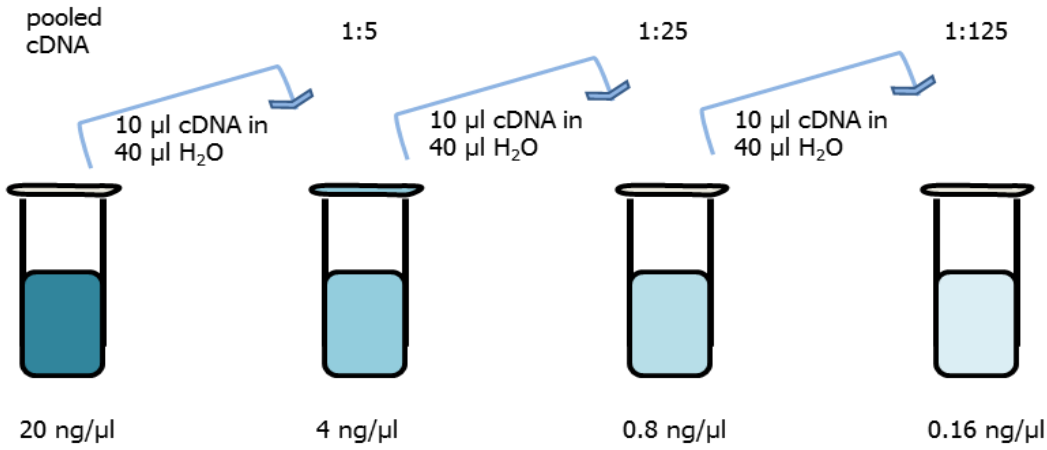

2.900 E-1
2.400 E-1
1.900 E-1
1.400 E-1
9.000 E-2
4.000 E-2
Fluorescence (-ΔF/ΔT)
NTC
60.4
65.0
70.0
75.0
80.0
85.0
90.0
95.0
Temperature °C

Figure 2.4 Principles of RT-qPCR. A) Diagrammatic illustration of SYBR Green chemistry in RT-qPCR. The reaction tube contains SYBR Green qPCR master mix supplied as a 2X pre-mixed formulation of Taq DNA polymerase, $MgCl_2$, dNTPs and Sybr Green I, supplemented with forward and reverse PCR primers and the cDNA template. During PCR cycling, DNA polymerase uses the primers with 3'-OH groups to incorporate dNTPs on single DNA strands and initiate double-strand synthesis. SYBR Green, a fluorescent dye with binding affinity for double-stranded molecules, will bind the DNA minor groove and enhance its fluorescence. As DNA polymerase elongates the cDNA template, more SYBR Green molecules will bind to the amplicon, producing an increased fluorescence signal. Eventually, PCR product will accumulate in successive cycles, resulting in exponential increase of the fluorescence signal. **B)** Real-time amplification curve with fluorescence plotted against PCR cycle. Dissociation curves for RT-qPCR products. In the early PCR cycles, a horizontal baseline is observed where fluorescence intensity is too low for detection above background. As the reaction proceeds, sample amplification proceeds exponentially, ideally doubling amplicon amounts at every cycle. The point at which the fluorescence crosses the detection threshold, which is set above the background within the exponential growth region of the amplification curve, is termed the threshold cycle, C_T, and is used for quantifying the input amount of a target in the original template. During the linear phase, reaction components become limited and amplicons are no longer doubled at each cycle. Eventually, reactions slow down and enter a plateau phase (end-point), where PCR reaction ceases. Real-time PCR using fluorescence-based detection measures PCR amplification as it occurs, whereas traditional PCR relies on limiting end-point analyses, which give ambiguous quantitation of the starting amounts of a target in the PCR reaction. **C)** During melting curve analysis, increasing temperature causes dsDNA to dissociate which in turn releases the intercalating SYBR Green dye and decreases the fluorescence signal. The change in fluorescence is plotted as a function of temperature to obtain a melt curve for the amplified product. Derivative melting curve plots of $-\Delta F/\Delta T$ (change in fluorescence/change in temperature) against temperature produce a melting peak profile. Every PCR product melts at a characteristic Tm, so that presence of a single peak suggests a single amplicon. Additional melting peaks could be a result of primer-dimer artifacts or non-specific amplicons. Signal detection in the water control sample (non-tempate control; NTC) would suggest contaminating template in the reaction components.

2.6.3 Construction of standard curves and relative quantification

Relative quantification is based on normalizing the expression of a target gene against an internal reference gene, expression of which is unaffected by the experimental condition, and on quantitating changes in target gene expression by calculating the ratio between the normalized amount of an experimental and a control (calibrator) sample. An external standard curve was constructed from cDNA samples expressing both the target and reference genes, as shown below:



Equal amounts of cDNA from the experimental and control samples analyzed on the reaction plate were pooled and serially diluted by taking 1/5 of the preceding standard and diluting it 5-fold with DNase-free water. Each dilution was assayed in triplicate. Given that 200 ng RNA was reverse-transcribed in a 10 µL reaction (20 ng/µL), the cDNA concentration of each dilution point on the standard curve is thus known and can be used to deduce the input amount of target template in experimental samples by comparing the C_T values of the unknown samples against the standards. To ensure that the serial dilution curve covered the expected range of target gene expression while accounting for varying C_T values between samples as a result of variation in starting target amounts, a 4-point 5-fold dilution standard curve was constructed and individual cDNA samples diluted 1:10 compared to the highest standard-curve concentration. Following qPCR amplification, the C_T data for each sample in comparison to the known standards was automatically determined using 7900 Fast System SDS 2.4 software (Applied Biosystems), and was exported for analysis in a Microsoft Excel spreadsheet. Before measuring

relative gene expression, the amplification efficiency of the PCR reaction was calculated from the slope of the standard curve using:

$$\text{efficiency} = 10^{(-1/\text{slope})}$$

Amplification efficiency was also converted into a percentage using:

$$\%\ \text{efficiency} = (10^{(-1/\text{slope})}) - 1)\ \text{x}100$$

An amplification efficiency of 100% would suggest a perfect 2-fold increase of amplicon per cycle during exponential phase of the reaction and would show evenly spaced amplification curves separated by 3.32 cycles for a 10-fold dilution (**Figure 2.5**). An acceptable PCR efficiency ranges between 0.9–1 (-3.6≥ slope ≥-3.3), with values outside that range suggesting suboptimal reaction conditions and co-amplification of nonspecific products, causing decreased sensitivity and detection artifacts, respectively.

Efficiency was also assessed using the R^2 value (the coefficient of determination; the square of Pearson's correlation coefficient) of the standard curve, which describes how linear the relationship is between the dilution series and their corresponding C_T, and serves as a quality measure of how close the experimental data are to the fitted regression line. For perfect correspondence, R^2 would equal 1.0, with deviations suggesting variability across standard-curve points, resulting in slope imprecision and inaccurate quantitation results. To enhance PCR efficiency and increase the R^2 value, dilution replicates with high C_T variations were excluded and dilution points at both ends of the standard curve producing high and low C_T outliers (replicate sample with a ΔC_T >0.3 from the mean) as a result of PCR inhibition and low target input, respectively, were deleted.

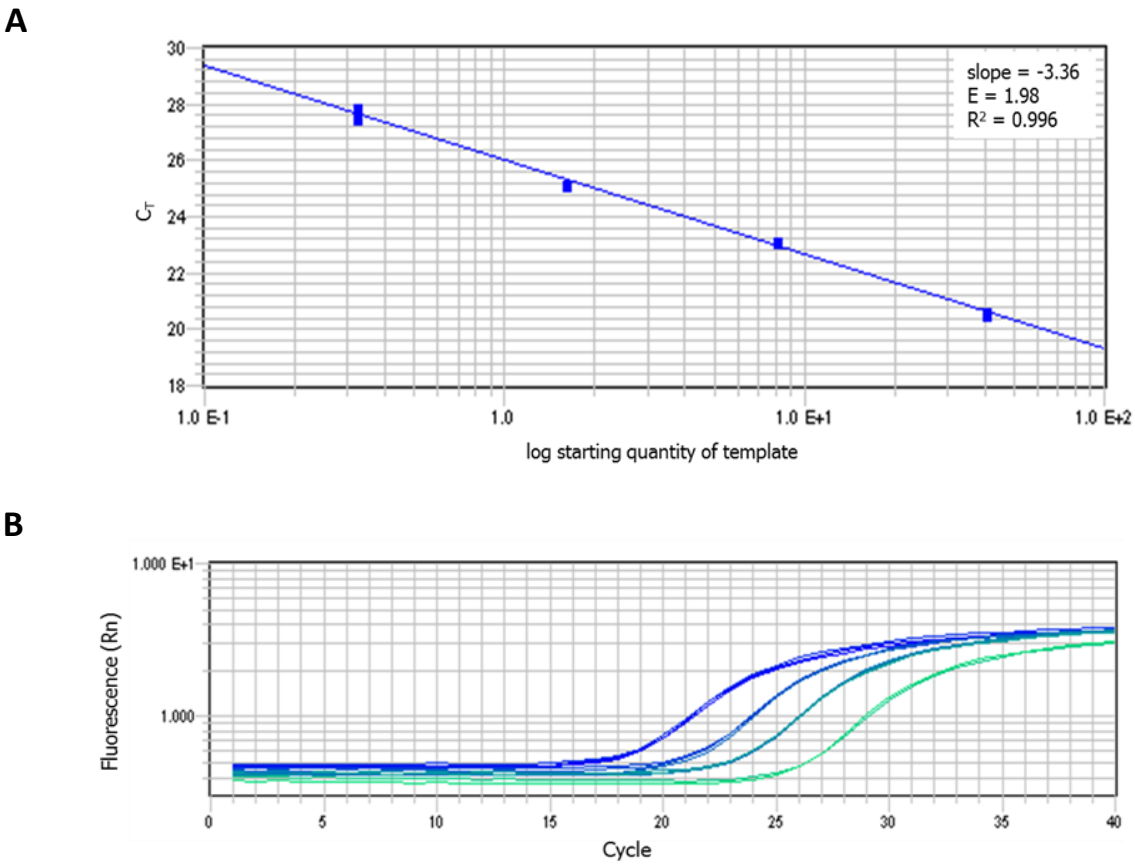


Figure 2.5. Standard curve construction to assess PCR reaction efficiency. A) Standard curve was constructed by plotting the C_T against the log of the starting quantity of the template for each dilution standard. The slope and R^2 values automatically generated by SDS 2.4 (*Applied Biosystems*) and the calculated PCR efficiency (E) are shown. **B)** Amplification curves of dilution standards, each run in triplicate. The curve and C_T values for replicates at each point are extremely close, suggesting low technical variability.

The amount of the target gene in the sample was normalized to the amount of the endogenous control GAPDH in the same sample using the following formula:

$$\text{normalized amount of target gene} = \frac{\text{target (ng)}}{\text{GAPDH (ng)}}$$

The relative expression of the target gene was quantitated by dividing the normalized target amount in the treated sample to the normalized target amount in the control sample:

$$\text{expression ratio (normalized to GAPDH)} = \frac{\text{treated sample (ng)}}{\text{control sample (ng)}}$$

Alternatively, the percent regulation was determined by dividing the difference in the normalized target amount between the treated and control samples to the normalized target amount in the control sample

$$\text{Expression ratio (normalized to GAPDH)} = \frac{\text{treated sample (ng)} - \text{control sample (ng)}}{\text{control sample (ng)}} \times 100$$

A paired *t*-test was used to compare mean expression ratios between treated and control samples, with a *p*-value ≤0.05 suggesting statistical significance. Descriptive data were presented as mean ± standard deviation (SD), where SD was computed using the standard deviation function of the *Excel* spreadsheet.

2.7 Flow cytometry analysis

2.7.1 Instrument

Flow cytometry was performed using a CyFlow Cube 8 v.6 channel instrument (*Sysmex*), equipped with a dual-laser system (488 nm blue & 632 nm red) and 6 optical parameters (Forward Scatter, Side Scatter and 4 fluorescence channels (FL1–FL4)) with the Windows™ based CyView Cube 1.3.2.4 software. Data analysis was performed with FCS Express 5 Flow Cytometry software (*De Novo Software, CA, U.S.A.*).

2.7.2 Sample preparation

Cell differentiation was assayed by flow-cytometric analysis of CD45, CD71 and CD235α expression levels. Cultured erythroblasts, ≥0.1x10^6, were re-suspended in 90 µL 2% FBS in PBS and blocked with FcR blocking reagent (10 µL/≤10^7 cells) for 10 minutes at 4 °C. Samples were subsequently stained with PerCP-CD45 (5 µL/10^6 cells), APC-CD71 (20 µL/10^6 cells), and PE-CD235α (0.02 µg/10^6 cells) on ice for 20 minutes in the dark. Cells were washed twice with 1 mL 2% FBS in PBS at 300 g for 5 minutes. Cells were then suspended in 500 µL 2% FBS in PBS and stored in the dark at 4 °C until analysis. A minimum of 10,000 live cell events were recorded. Mean fluorescence intensity (MFI) was used as a measure for protein expression.

FcR blocking reagent; 130-059-901	*Miltenyi Biotecy, Germany*
PerCP-CD45; 304026 APC-CD71; 334108 PE-CD235α; 306604 (GpA) PE-$CD34^+$; 343606	*Biolegend, CA, USA*
PE-CD117 (c-kit); 12-1178-42	*eBioscience, Affymetrix, CA, USA*

2.8 Mammalian cell cultures

2.8.1 Cell lines and culture conditions

The following cell lines were grown and maintained for experimental use:

Cell line	Origin	Ref.
HEL	Human erythroid leukemia cells	Martin, P., 2001 [250]
K562	Human erythromyeloblastoid leukemia cells	Rutherford, T., et al, 1981 [251]
MEL	Murine pro-erythroid leukemia cells	Levenson, R., & Housman, D., 1979 [252]
HEK 293T	Human embryonic kidney cells	Graham, F.L., & Smiley, J., 1977 [253]

HEL and K562 cells were maintained in RPMI 1640, MEL cells were maintained in DMEM, and HEK 293T cells were maintained in IMDM. Media were supplemented with 10% fetal bovine serum (FBS), 1% penicillin-streptomycin (P/S) and 1% glutamine (all *Invitrogen*^{TM}) (complete media; e.g.: cRPMI). Cells were incubated at 37 °C in 5% CO_2 humidified atmosphere.

Cells were passaged upon reaching 80%–90% confluence, usually twice a week. Non-adherent cells (HEL, K562 and MEL) were maintained by culturing a proportion of cells at a density of $0.2x10^6$ cells/mL in fresh medium. Adherent cells (HEK 293T) were maintained at 1/10 dilution in 20 mL fresh medium in 75-cm^2 flasks (*Corning, NY, USA*) by adding 2 mL 1X Trypsin-EDTA (*Invitrogen*^{TM}) followed by incubation at 37 ˚C for 2 minutes and then diluting trypsinized cell suspension with growth medium.

2.8.1.1 Trypan blue exclusion assay

Living cell counts were determined using trypan blue assay [254] by adding equal volumes of cell suspension to 0.4% trypan blue stain (*Sigma-Aldrich*) and scoring for blue-stained apoptotic and necrotic cells on a Bright-Line™ haemocytometer (*Sigma-Aldrich*) under a light microscope (Eclipse TS100, *Nikon, Chiyoda, Tokyo, Japan*). Alternatively, cells were counted using the Countess™ automated cell counter (*Invitrogen™*).

2.8.1.2 Hoechst 33258 DNA staining for Mycoplasma

Hoechst 33258 stock solution was prepared by dissolving 5 mg powder (*Sigma-Aldrich)* in 10 mL PBS and was further diluted at a ratio of $1{:}10^4$ (µL, v/v) in PBS before use. A 2 mL cell suspension was mixed with 50 µL of diluted DNA stain and incubated at RTemp for 15 minutes, followed by visual examination for extranuclear fluorescence under a fluorescent microscope with a UV filtration package (Eclipse TE2000-U; *Nikon*).

2.8.2 Production of transgenic cell lines

Cells ($2x10^5$) were cultured in 0.5 mL cRMPI 1680 medium supplemented with 8 µg/mL polybrene (*Sigma-Aldrich*) in 24-well plates (*Corning costar*) and transduced with vectors of titers $>10^8$ transduction units (TU)/mL at a multiplicity of infection (MOI) 5. The cell-virus mix was incubated at 37 °C in 5% CO_2, overnight, followed by seeding at limiting dilution in 96-well plates (*Iwaki, Japan*) in the presence of conditioned medium (*) to support cell viability and growth. Wells with single cells were marked, expanded and subcultured into larger vessels. Cells were incubated at 37 °C in 5% CO_2 humidified atmosphere.

(*) The conditioned medium was derived from the HEL cell line. The growth medium of confluent cultures was collected and centrifuged at 300 g for 5 minutes. The supernatant was filter-sterilised and stored in 15-mL aliquots at -20 °C.

2.8.3 Induction of erythroid differentiation in cell lines

2.8.3.1 Comparative analysis of erythroid inducers in HEL cells

Phosphoglycerate kinase 1 (PGK)-GFP-expressing HEL cells were cultured in cRPMI 1640 medium in 10-cm plates (*Greiner Bio-One, NC, USA*) at a concentration of $0.2x10^6$ cells/mL in the presence of the following concentrations of known erythroid inducers: 0.5% dimethylsulfoxide (DMSO) (*Sigma-Aldrich;* D8418), 10 U/mL erythropoietin (EPO) (*Binocrit, Novartis*), 0.5 mM sodium butyrate (*Sigma-Aldrich;* B5887*),* and 30 µM, 50 µM and 100 µM haemin (*Sigma-Aldrich;* 51280), respectively. DMSO solution was stored at RTemp. EPO was stored at 4 °C. Sodium butyrate stock solution was prepared in dH_2O at 1 M, filtered through a 0.22-*µ*m filter and stored at -20 °C. Haemin stock solution was prepared in neat DMSO at 53 mM, filtered through a 0.45-*µ*m filter and stored at -20 °C. Stock solutions were diluted in medium before addition to the cultures. Treatment with inducers was repeated on day 2 of culture, and cells were collected on day 4 of culture. Each treatment was carried out in triplicate, in parallel to un-induced control cultures. Cells were incubated at 37 °C in 5% CO_2 humidified atmosphere. Cell differentiation was assayed using the benzidine test, and cells were screened for surface differentiation marker expression, namely glycophorin A (GpA) and transferrin receptor (CD71) by flow cytometry. Cell viability was assessed using the trypan blue exclusion.

2.8.3.2 Comparative analysis of DMSO concentrations in MEL cells

βLCR-GFP-expressing MEL cells with inducible GFP expression were cultured at a concentration of $1.5x10^5$ cells/mL in 2 mL cDMEM medium in a 24-well plate (*Corning costar*). Differentiation was assessed using increasing concentrations of DMSO (*Sigma-Aldrich*) at 0.5%, 1%, 1.5% and 2%, each carried out in triplicate. Control cultures were grown in cDMEM supplemented with increasing concentrations of FBS (*Invitrogen*^{TM}) at 5%, 10% and 20%, again in triplicate. Cells were incubated at 37 °C in 5% CO_2 humidified atmosphere. On day 4 of the culture, cells were collected for assessment of differentiation using benzidine staining, of DMSO-induced toxicity using trypan blue exclusion assay and of GFP expression levels using flow cytometry.

2.8.3.3 Erythroid differentiation of K562 cells

K562 cells were differentiated in the presence of 100 μM hydroxyurea (*Sigma-Aldrich;* H8627). A stock solution of 1 M was prepared by dissolving hydroxyurea powder in sterile water and stored at -80 °C. Tube caps were wrapped with parafilm. Working solutions of 10 mM hydroxyurea were prepared by dilution in sterile water and were further diluted to 100 μM in medium before addition to the cultures.

2.8.4 Transduction of cell lines

Cells were cultured at a concentration of 0.2–0.5x10^6 cells in 0.5 mL growth medium supplemented with 8 μg/mL polybrene (*Sigma-Aldrich*) and were transduced with vectors of >10^8 TU/mL at an MOI 5–50 (details specified in text). Transductions were performed in 24-well plates over a 6-hour incubation of the cell-virus mix at 37 °C with hourly resuspension. After the 6-hour incubation, 1 mL growth medium was added to each well and cells were returned at 37 °C. 24 hours later, cells were collected and cultured in 4 mL growth medium supplemented with an appropriate inducing agent in 25 cm^2 Nunc™ treated EasYFlasks (*Thermo Scientific*). Cells were incubated at 37 °C in 5% CO_2 humidified atmosphere. Fresh inducer-supplemented growth medium (4–5 mL) was added on days 3 and 5. Cells were collected on day 7.

2.9 *In vitro* models of human erythropoiesis

2.9.1 Study samples

Buffy coats from normal donors (50–70 mL) were obtained from the National Blood Bank in Nicosia. PB (15–20 mL) from thalassaemic donors was provided by Dr. Maria Sitarou from the Thalassaemia Department, Larnaca General Hospital, Cyprus. Small-volume pre-transfusion thalassaemic blood samples (2–4 mL) were obtained from the Cyprus Thalassaemia Centre, Nicosia. All subjects gave written informed consent. Experimental protocols were approved by the Cyprus National Bioethics Committee (Applications ΕΕΒΚ/ΕΠ /2012/02 “Advancing Gene Therapy Vectors for Thalassaemia” and ΕΕΒΚ/ΕΠ /2013/23 “ThalaMoSS”).

2.9.2 Peripheral blood cell extraction

2.9.2.1 Isolation of mononuclear cells from peripheral blood using Ficoll density gradient centrifugation

PB was collected into EDTA-coated tubes (*Greiner Bio-One*; 455036). Both the buffy coat and PB were mixed with 1.5 volumes PBS (without Ca^{2+}/Mg^{2+}) to reduce erythrocyte clumping. In 50-mL tubes, 1.25 blood-volumes of RTemp density gradient reagent (either Ficoll Paque^{TM} PLUS, *GE Healthcare Life Sciences, PA, U.S.A.*, or Lymphoprep, *Accu-Prep^{TM} Lymphocytes*, *Axis-Shield PoC AS, Norway*), density 1.077 g/mL, was added and gently overlaid with 2 volumes of blood-PBS mix. The PB mononuclear cell (PBMC) fraction was separated from blood by Ficoll density gradient, **Figure 2.6**, via centrifugation at 900 g, brake off, for 25 minutes at RTemp in a swing-out rotor, extending to 35 minutes for blood samples collected ≥2 hours before procedure. Using a 5-mL pipette, the visible ring of mononuclear cells from the interphase was collected in 50-mL tubes. In cases where blood fractionated under the same conditions resulted in PBMC layer contamination with nucleated red cell precursors, a second round of Ficoll density gradient separation was performed. The PBMC layer was washed, twice, in PBS at 300 g for 5 minutes at RTemp, followed by a third wash at 200 g for 5 minutes at RTemp to eliminate platelet contamination. During washes, the supernatant was discarded and the cell pellet was disrupted by scraping the tube against a ribbed surface. The cell pellet was resuspended in 600 µL of cold 1% bovine serum albumin (BSA) (*Calbiochem, CA, USA*) in PBS.

2.9.2.2 Isolation of PB nucleated cells using ACK lysis buffer

Samples of PB collected into EDTA anticoagulant (*USB*) were mixed with ACK lysis buffer (0.15 M NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA, pH 7.2) at 1:2 volume ratio by gentle inversion, incubated at RTemp for 10 minutes and centrifuged at 200 g for 10 minutes. The cell pellet was suspended with 1 blood-volume ACK lysis buffer and centrifuged at 200 g for 10 minutes at RTemp. The resulting nucleated cell fraction was re-suspended in 1 mL of 2% FBS in PBS supplemented with 2 mM EDTA.

2.9.2.3 Isolation of PBMCs using Ficoll-Paque PLUS with ACK lysis buffer

Adapted from [255], briefly, blood was washed twice with RPMI at 200 g for 10 minutes at RTemp, and the resulting pellet was treated with ACK lysis as described in **section 2.9.2.2**. Nucleated cells (NCs) were resuspended in RPMI and layered over Ficoll-Paque^{TM} PLUS (*Amersham Biosciences, NJ, USA*) following density gradient separation as described in **section 2.9.2.1**.

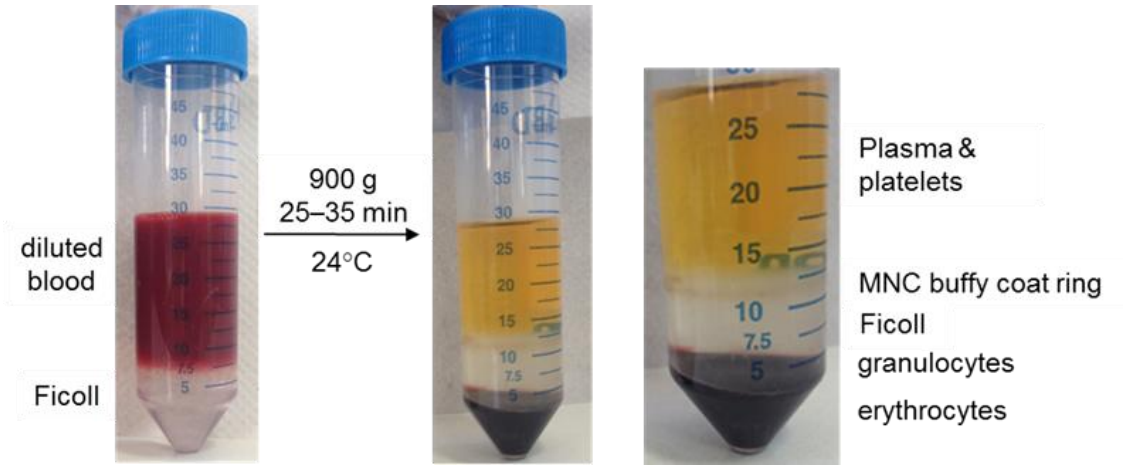


Figure 2.6. PB mononuclear cell isolation using Ficoll gradient density centrifugation.

2.9.2.4 Enrichment of CD34^{+} cells using Magnetic Activating Cell Sorting

PB-derived cells (**sections 2.9.2.1–2.9.2.2**) were incubated with 100 µL anti-human CD34 MicroBeads (*Miltenyi Biotecy, Bergisch Gladbach, Germany*) for 15 minutes on ice on a rocking platform in the cold-room. Beading buffer (0.5% BSA, 2 mM EDTA in PBS (without Ca^{2+}/Mg^{2+})) was added to the cell-bead mix to fill up the tube, followed by centrifugation at 300 g for 5 minutes at 4 °C with brake on and maximum acceleration. Washing with beading buffer was repeated once, and cells were re-suspended in 1 mL ice-cold buffer. Magnetic separation was carried out with LS MACS columns (*Miltenyi Biotecy*). The MidiMACS separator and Multistand (*Miltenyi Biotecy*) were pre-chilled by incubation for ≥3 hours at 4 °C, and LS columns (2 columns/sample) were equilibrated with 5 mL of ice-cold beading buffer. Cell suspension, 0.5

mL, was applied to the column followed by washing with 1.5 mL beading buffer. This step was repeated to load the remaining cells. The column was washed twice with 3 mL of ice-cold beading-buffer. The previous buffer was allowed to elute completely before each subsequent step, and all reagents were added at the same pace. The column was removed from the magnetic separator and placed as quickly as possible on top of a clean 15-mL tube. Beading buffer, 5 mL, was added quickly and cells were flushed by inserting the plunger into the column. The second pre-chilled and equilibrated column was loaded on the magnetic separator. The selection process was repeated by adding 2.5 mL of first-round-selected cells followed by 2.5 mL beading buffer, until all cells were applied on the column. The column was washed twice, with 3 mL beading buffer, and cells were eluted in 5 mL ice-cold buffer. $CD34^+$-enriched cells were collected at 300 g for 5 minutes at RTemp.

2.9.3 Freezing of cells isolated from PB

PB-derived cells (**sections 2.9.2.1–2.9.2.2**) were frozen upon extraction in 1 mL freezing medium, distributed to cryovials on ice and transferred to a controlled-rate freezing isopropyl-containing Mr. Frosty container (*Nalgene, NY, USA*) for -80 °C storage. Two freezing media were prepared and used as indicated. Freezing media A and B consisted of 10% DMSO in IMDM supplemented with either 30% FBS (*Invitrogen*^{TM}; 10500) or 50% characterized FBS (Hyclone SH30071.03, *Thermo Scientific*), respectively. The freezing media were prepared fresh and chilled on ice before use. $CD34^+$-enriched cells (**section 2.9.2.4**) were expanded in culture (**section 2.9.5.2.1**) for 10 days, collected at 300 g for 5 minutes and re-suspended in 1 mL freezing medium B at a concentration of $\geq 1x10^6$ cells/mL. Cells were aliquoted in chilled cryovials and transferred from ice to a cotton-wool-insulated pre-cooled box at -80 °C. Cryovials were transferred to liquid nitrogen within 30 days of freezing for long-term storage.

2.9.4 Thawing of cells isolated from peripheral blood

Cryotubes of frozen blood-derived cells were placed in 37 °C water, and thawed cells were added, drop by drop, to 10 mL defrosting medium containing 5% characterized FBS in IMDM. Solutions were incubated at RTemp for 5 minutes, and cells were collected by centrifugation at 200 g for 10 minutes at RTemp. Cells were re-suspended in growth medium and scored for

viability using trypan blue. Cells were either seeded in semi-solid methylcellulose cultures (**section 2.9.6**) or cultured in fresh expansion medium at a concentration of 0.5–1x10^6 cells/mL in 6-well Nunc plates (*Thermo Schientific*).

2.9.5 Erythroid liquid cultures

2.9.5.1 Erythroid liquid culture – Protocol A

2.9.5.1.1 Pre-stimulation of $CD34^+$ cells

Progenitor cells were isolated from 40–70 mL healthy donor buffy-coat samples using either Ficoll density gradient (**section 2.9.2.1**) or ACK lysis buffer (**section 2.9.2.2**), followed by positive selection using anti-CD34 magnetic beads and MACS columns according to manufacturer's instructions (*Miltenyi Biotecy*). $CD34^+$ cells were counted and plated at a concentration of 0.5-1x10^6 cells/mL in pre-activation medium consisting of Cell Gro SCGM (*Cell Genix GmbH*, *Freiburg, Germany*) supplemented with 300 ng/mL human stem cell factor (SCF) (*Peprotech. NJ, USA*; 300-07), 300 ng/mL human Flt3-ligand (Flt3L) (*Peprotech*; 300-19), 100 ng/mL human thrombopoietin (THPO) (*Invitrogen*; PHC9514) and 60 ng/mL human interleukin-3 (IL3) (*Peprotech*; 200-03) in 35-mm plates (*Corning*; 430588).

2.9.5.1.2 Transduction of $CD34^+$-enriched cultures

Transduction of $CD34^+$ cells was performed on retronectin-coated plates prepared according to manufacturer's instructions. Briefly, retronectin (*Takara, Kyoto, Japan*) was dissolved in 1 mL PBS at a concentration of 43 µg/mL and spread on the surface of a 35-mm plate (5 µg/cm^2) for 2 hours at RTemp. Then, retronectin solution was removed and replaced with 1 mL 2% BSA in PBS for 30 minutes at RTemp, following a subsequent wash with 1 mL PBS. $CD34^+$ cells were plated on retronectin-coated plates in pre-stimulation medium as described in **section 2.9.5.1.1**. 24 hours later, cells were transduced overnight at a MOI of 100, unless otherwise stated, and at a viral concentration of >10^8 TU/mL in the presence of 4µg/mL polybrene.

2.9.5.1.3 $CD34^+$ expansion and differentiation

As adapted from [188], $CD34^+$ cells were cultured in cIMDM medium containing 20% FBS (*Invitrogen*^{TM}; 10500), 10 ng/mL SCF (*Peprotech,* 300-07), 1 ng/mL IL3 (*Peprotech*; 200-03), 1 U/mL EPO (Eprex 4000), 10^{-6} M dexamethasone (*Sigma-Aldrich*) and 10^{-6} M β-estradiol (*Sigma-Aldrich*) at a concentration of $0.5x10^6$/mL in 35-mm plates (*Corning*; #430588). Cells were cultured for 12–14 days at 37 °C in 5% CO_2 humidified atmosphere and were maintained at $1x10^6$/mL by counting and re-culturing in fresh medium.

2.9.5.2 Erythroid liquid culture – Protocol B

2.9.5.2.1 Erythroid expansion and selection of HSPCs (Phase-I)

PBMCs were isolated by density gradient centrifugation and cultured in 40 mL Phase I medium consisting of α-MEM (*Sigma-Aldrich*; M0644) supplemented with 10% FBS (*Invitrogen*^{TM}; 16000), 1% glutamine (*Invitrogen*^{TM}), 5 U/mL P/S (*Invitrogen*^{TM}), 10 ng/mL SCF (*Invitrogen*^{TM}), 1 µg/mL cyclosporin A (*Sigma-Aldrich*), and 10% conditioned medium from cultures of H5637 bladder carcinoma cell line. Cells were cultured in 75 cm^2 Nunclon^{TM} Delta flasks (Nunc; *Thermo Scientific*) for 7 days at 37 °C in 5% CO_2 humified atmosphere.

All reagents were prepared and provided by Andria Theodorou, CING, Nicosia, Cyprus.

2.9.5.2.2 Erythroid differentiation of HSPCs (Phase-II)

At the end of Phase-I culture, non-adherent cells were harvested and centrifuged at 266 g for 8 minutes. The cells were washed with 1X PBS at 266 g for 8 minutes, and the pellet was re-suspended in 40 mL Phase-II medium consisting of α-MEM (*Sigma-Aldrich*), 30% FBS (*Invitrogen*^{TM}), 0.01 mM β-mercaptoethanol (*Sigma-Aldrich*), 0.77 µM dexamethasone (*Sigma-Aldrich*), 1 U/mL recombinant EPO (*Janssen-Cilag Lt, Bucks, UK*), 10 ng/mL SCF (*Invitrogen*^{TM}), 1% BSA (*Sigma-Aldrich*), 5 U/mL P/S (*Invitrogen*^{TM}), and 2 mM glutamine (*Invitrogen*^{TM}). Cells were cultured in 75 cm^2 Nunclon^{TM} Delta flasks (*Nunc*) for 10–14 days at 37 °C in 5% CO_2 humidified atmosphere.

2.9.5.3 Erythroid liquid culture: Protocol C

2.9.5.3.1 Expansion of $CD34^+$ cells (Phase-I)

$CD34^+$ cells were isolated from PBMCs using density gradient centrifugation and were cultured in 1.5–3 mL expansion medium consisting of StemSpan SFEM II (*Stemcell Technologies, Vancouver, Canada*; 09650) supplemented with 1X CC-100 (*Stemcell Technologies*; 02690), 2 U/mL EPO (*Sandoz GmbH*, Austria, Binocrit 4,000 IU/0.4 mL), 10^{-6} M dexamethasone (*Sigma-Aldrich*; D4902) and 1X P/S (*Corning-Cellgro*; 30-002-CI). Cells were cultured in 12-well plates (*BD Falcon*; 353043) or 6-well Nunc plates (*Thermo Scientific*; 140685) and were propagated at a concentration of $0.25x10^6$/mL at least twice a week. Cultures of >5 mL volume were grown in 25 cm^2 Nunc flasks (*Thermo Scientific*; 156367). Cells were expanded in culture for 10–14 days at 37 °C in 5% CO_2 humidified atmosphere.

2.9.5.3.2 Differentiation of $CD34^+$ cells (Phase-II)

$CD34^+$ cells were cultured in differentiation medium containing 70 % αMEM (*Corning CellGro*; 10-22-CV), 30 % defined FBS (*Hyclone*; SH30070.03) , 10^{-5} M 2-mercapto-ethanol (*Sigma-Aldrich*; M3148), 10 U/mL EPO (Binocrit 4,000, *Sandoz GmbH*) and 10 ng/mL SCF (*Peprotech*; 300-07) at a concentration of $0.5–1x10^6$/mL in 6-well Nunc plates (*Thermo Scientific*; 140685). Cultures were diluted over time with fresh medium to maintain the cell concentration at $1x10^6$/mL. Cells were collected on day 6 or 8.

2.9.5.3.3 Transduction of $CD34^+$ cells

$CD34^+$ cells ($1–2x10^6$) were suspended in 0.5 mL expansion medium supplemented with 8 µg/mL polybrene (*Millipore, Massachusetts, USA*) in sterile 1.5 mL microcentrifuge tubes. Vectors of $>1x10^8$ TU/mL were added to achieve an estimated VCN of 0.2–2 (with an MOI in the range of 0.5–5) and the cell-virus mixture was incubated for 6 hours at 37 °C, during which time the solution was mixed on an hourly basis by pipetting up and down. After the 6-hour incubation, cells were collected at 200 g for 5 minutes at RTemp and resuspended in expansion medium (**section 2.9.5.3.1**) at a concentration of $0.5–1x10^6$/mL in 6-well Nunc plates for 2–3 days. Thereafter, cells were collected at 200 g for 5 minutes and sub-cultured in differentiation

medium at a concentration of $1x10^6$/mL. Fresh differentiation medium (1–3 mL) was added, dropwise, after 3 days, and cells were collected on day 6–8.

2.9.6 Colony forming cell (CFC) assay using methylcellulose-based media

Blood-derived cells (MNCs, NCs) were counted by trypan blue staining and adjusted to cell plating concentrations of $0.15x10^6$ cells per 35-mm plate (*Corning*; 430588) in 1.1–1.5-mL methylcellulose complete media (*R&D systems, Minneapolis, USA*; HSC003) according to manufacturer's instructions. Sample plates were incubated at 37 °C in 5% CO_2 for 14 days, followed by colony scoring using an AxioVert 200 inverted microscope with fluorescence imaging (*Carl Zeiss AG, Oberkochen, Germany*) to enumerate and quantify multi-potent and single lineage haematopoietic progenitors based on their morphological characteristics (**Figure 2.7**).

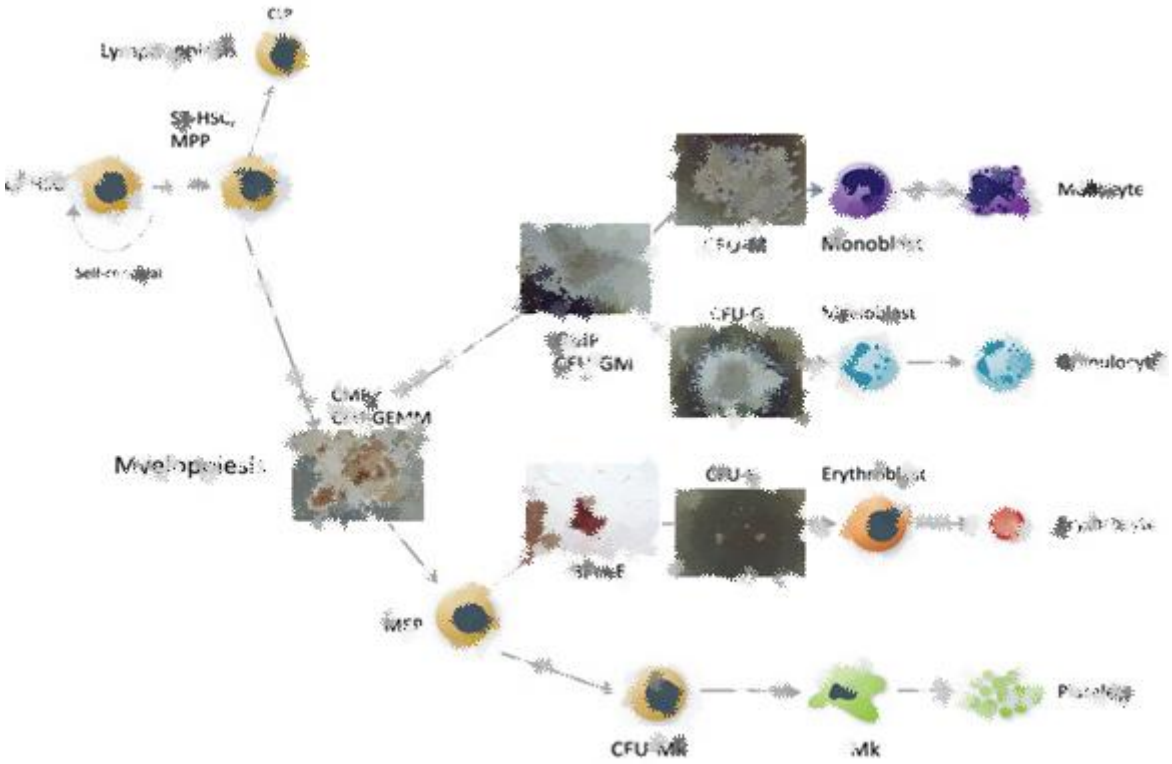


Figure 2.7. Schematic of haematopoietic development directed towards myeloid differentiation using cytokine-formulated semi-solid methylcellulose. LT-/ST-HSC: long- and short-term-repopulating haematopoietic stem cells, respectively; MPP: multipotent progenitors; CLP: common lymphoid progenitors; CMP: common myeloid progenitors; CFU-GEMM: colony forming unit-granulocyte, erythrocyte, macrophage, megakaryocyte; GMP: granulocyte-monocyte progenitor; CFU-GM: colony forming unit-granulocyte, macrophage; CFU-M: colony forming-unit macrophage; CFU-G: colony forming-unit granulocyte; MEP: megakaryocyte-erythrocyte progenitor; BFU-E: burst forming unit-erythroid; CFU-E: colony forming unit-erythroid. Adapted from [256, 257].

2.10 Assessment of erythroid differentiation

2.10.1 Benzidine staining

Erythroid differentiated cell lines and primary erythroblasts were scored for haemoglobin production by benzidine assay. Benzidine stock solution was prepared by dissolving 2 mg/mL dihydrochloride benzidine (*Sigma-Aldrich*) in 3% acetic acid and stored at 4 °C. Equal volume of benzidine working solution, prepared by adding 20 µL 30% hydrogen peroxide (*Merck*) to 1 mL stock solution, was added to cells followed by incubation on ice for 2 minutes. Benzidine-stained cells were counted using a Bright-Line^{TM} haemocytometer (*Sigma-Aldrich*) under a light microscope (Eclipse TS100; *Nikon*).

2.10.2 Cytospin preparation and May-Grünwald-Giemsa staining

Phenotypic characterization of erythroid progenitors in primary cultures was based on cell morphology using cytospin preparations. Suspensions of $0.6x10^6$ cells per 0.2 mL medium were deposited on slides at 95 g for 3 minutes (acceleration: 6, break: 6) using the Tharmac Cellspin II cytocentrifuge with an EASY rotor (*Hettich*, *Germany;* A320). Smears were air-dried and slides were stained in pure May-Grünwald solution (*Sigma-Aldrich*) for 5 minutes at RTemp, rinsed in water and subsequently stained in pure Wright-Giemsa stain buffer (*Fluka Analytica*, *Sigma-Aldrich*) for 10 minutes at RTemp. The slides were washed in water, air-dried, covered with mounting medium (*Dakocytomation, CA, USA*) and sealed with a coverslip. Cytospins were viewed under the Nikon Eclipse TE2000-U miscroscope equipped with a DN100 digital network camera for image capture (*Nikon*).

2.11 DNA cloning

2.11.1 DNA digestion by restriction endonucleases

A plasmid vector and a target DNA (1–5 µg) were digested with selected restriction endonuclease enzyme(s) (*NEB*®) to produce compatible restriction ends according to manufacturer's instructions. Reactions were performed in the presence of compatible 10X NEBuffers (*NEB*®) complemented with supplied 10X BSA (*NEB*®) and two units of enzyme (varied according to enzyme efficiency) which constituted <10% of the reaction volume to prevent star activity. Digestions were performed over 3 hours to overnight incubation at recommended temperatures. Double enzyme digests with incompatible buffers were performed sequentially with an intermediate phenol-chloroform extraction and alcohol precipitation step (**section 2.1.1.2**) or by adjusting the salt concentration of the reaction, such that digestions in low-salt buffers were performed first, followed by addition of the second enzyme and a volume of its higher-salt buffer that would approximate its optimal reaction conditions.

2.11.2 Blunting ends using T4 DNA polymerase

T4 DNA polymerase (*NEB*®) was used for blunting ends in digested DNA by 3'-overhang removal or 5'-overhang fill-in as described by the manufacturer. Briefly, DNA was purified from restriction enzymes and buffers by phenol:chloroform:isoamyl extraction and alcohol precipitation (**section 2.1.1.2**) before addition of T4 DNA polymerase (1 U enzyme per 1 µg DNA) in a final reaction volume of 20 µL containing 2X NEBuffer 2 (*NEB*®) and 100 µM dNTPs. The reaction was incubated at 12 °C for 15 minutes, supplemented with 10 mM EDTA and heat-inactivated at 75 °C for 20 minutes.

2.11.3 De-phosphorylation of vector backbone

Vector DNA was dephosphorylated by adding 1 µL Antarctic phosphatase (*NEB*®) in the digestion reaction supplemented with 1/10 volume of supplied 10X Antarctic phosphatase reaction buffer. Reactions were incubated at 37 °C for 15 minutes for 5' extensions or blunt-ends, extending to 60 minutes for 3' extensions. The enzyme was heat inactivated at 65 °C for 30 minutes. DNA was precipitated with 70% isopropanol supplemented with 1/10 volume

sodium acetate pH 5.2 at 16,100 g (maximum speed) for 15 minutes, following a wash with 75% ethanol and final re-suspension in 10 µL dH_2O.

2.11.4 Extraction of DNA fragments

Digested DNA was separated on agarose gel for extraction of fragments of interest with the NucleoSpin® Gel and PCR Clean-Up Kit (*Macherey-Nagel*) according to manufacturer's instructions. Briefly, DNA-containing gel bands were lysed and washed through a Si-OH silica-based membrane column. DNA was eluted with 30 µL 5 mM Tris/HCl, pH8.5 buffer.

2.11.5 Ligation of DNA fragments

Vector and insert fragments were ligated at 1:3 molar ratio using either a standard ligation reaction [258] or a Quick Ligation™ Kit following manufacturer's protocol (*NEB*®). The amount of insert needed for ligation was calculated as follows:

$$\text{insert (ng)} = \frac{\text{vector (ng) x insert size (kb)}}{\text{vector size (kb)}} \times \frac{1}{3}$$

The DNA fragments separated on gels were visualized with the Vilber Lourmat UV transilluminator, and images captured with the Mega-Capt software. Band intensities were assumed as being proportional to DNA amounts and analysed by ImageJ with the MRI Cell Image Analyzer program (*Research Services Branch*, *NIH*) to calculate intensity ratios between DNA bands, which, together with expected fragment sizes (kb) allowed calculation of DNA amounts required for a 1:3 vector:insert molar ratio and a targeted vector amount of 50 ng per reaction. Ligation was performed with 1 µL T4 DNA ligase (*NEB*®) and supplied 10X T4 DNA ligase reaction buffer (*NEB*®) in a final reaction volume of 20 µL, followed by overnight incubation at 16 °C. Alternatively, 20 µL reactions with T4 DNA ligase (*NEB*®) and 2X Quick ligation buffer (*NEB*®) were incubated at RTemp for 5 minutes and chilled on ice pending immediate bacterial transformation.

2.11.6 Cloning oligonucleotides into pLKO.1 vector

Oligonucleoties were designed according to published guidelines of the RNAi consortium library (http://www.broadinstitute.org/rnai/public/resources/protocols; [259]) for expression of shRNAs under RNA-pol-III promoters, and were synthesized as independent complementary forward and reverse strands by *Metabion*. Annealing was performed in a final reaction volume of 50 µL using 5 µL of 20 µM forward and reverse oligonucleoties and 5 µL of 10X NEBuffer 2 (*NEB*®) on the Veriti® thermal cycler (*Invitrogen*™) programmed to start at 95 °C for 4 minutes for complete strand dissociation. Temperature was first lowered at a 10% maximum ramp rate to 80 °C for 4 minutes and then at 0.3% maximum ramp rate to 30 °C for 4 minutes to ensure proper annealing and generation of doube-stranded oligonucleotide inserts carrying overhangs compatible with *Eco*RI (5') and *Age*I (3') restriction sites. Annealed oligos (2 µL out of the 50 µL reaction) were ligated with 20 ng of correspondingly digested pLKO.1 vector, before 2 µL of the ligation reaction were transformed into 25 µL chemically competent *Escherichia coli* (*E.coli*).

2.12 Bacterial cultures

2.12.1 *E. coli* strains

Commercially available XL10-Gold® Ultracompetent Cells (*Stratagene, CA, USA*), DH5α™ and One Shot® TOP10 competent cells (*Invitrogen*™) and Turbo competent cells (*NEB*®) were used. Alternatively, frozen XL1 Blue *E.coli* competent cells were prepared as described in [260]. Briefly, 2 mL of a bacterial culture was sub-cultured in 400 mL antibiotic-selective broth, grown at 37 °C, and monitored by periodically reading the optical density on UV-1700 Pharma Spec (*Shimadzu*). At an OD_{600nm} of 0.48, the bacterial culture was incubated on ice for 30 minutes. The bacterial pellet was collected at 3,000 g for 10 minutes, suspended in 50 mL cold 0.1 M $CaCl_2$, and incubated overnight on ice in the cold-room. Then, 2 mL cold 100% glycerol was added to the bacterial suspension, followed by preparation of 50–100 µL aliquots which were snap-frozen in liquid nitrogen prior to storage at -80 °C.

2.12.2 Preparation of bacterial growth media

Bacterial growth media were prepared according to [261]. Briefly, lysogeny broth (LB) was prepared by dissolving 10 g/L tryptone, 5 g/L yeast extract and 10 g/L NaCl in dH_2O. Solid LB growth media containing bacto-agar (15 g/L) were prepared according to [262]. All LB media were sterilized by autoclaving. Antibiotic selection was conferred by supplementing media with 50–100 µg/mL ampicillin (*Sigma-Aldrich*) or kanamycin (*USB*), for liquid media immediately before use, for agar plates once autoclaved media had cooled to 50 ˚C and before pouring them into 10-cm plates (*Greiner Bio one*).

2.12.3 Bacterial transformation

Competent cells were thawed on ice for 10 minutes and mixed gently with plasmid DNA (20 ng in a volume of 25 µL cells) or with ligated DNA (1/10 the volume of 50 µL cells) followed by incubation on ice for 30 minutes. Heat shock was performed at 42 ˚C for 45 minutes and a subsequent incubation on ice for 5 minutes. Then, 1 mL broth medium without antibiotic was added, followed by incubation for 40–60 minutes at 37 ˚C with occasional mixing. Bacteria cells were collected at 5,000 g for 2 minutes, re-suspended in 100 µL LB and spread onto pre-warmed antibiotic selection plates. Alternatively, a quick transformation protocol was carried out as decribed in [263]. Briefly and for ampicillin selection, only, the mixture of competent cells and DNA was incubated on ice for 5 minutes followed by heat shock through immediate spreading on 37 ˚C pre-warmed agar plates. Transformation plates were incubated overnight at 37 ˚C.

2.12.4 Growth of bacterial cells and plasmid preparation

Single colonies of bacteria grown on agar plates were cultured in broth media containing 50–100 µg/mL antibiotic, followed by overnight growth at 37 ˚C with shaking at 250 rpm. For transformed ligated DNA, a 5-mL culture was prepared for standard mini preparations using phenol:chloroform plasmid extraction (**section 2.1.2.1**) and confirmation of cloning by diagnostic digests (**section 2.11.1**). Mini cultures of confirmed plasmids were inoculated in 1 L media at 1:500 for high DNA yield, and plasmid was purified using *Macherey Nagel* plasmid purification kits as described in **section 2.1.2.2**.

2.12.5 Making *E.coli* stocks

E. coli stocks were stored in growth medium diluted 1:1 (v/v) with 40% glycerol, followed by mixing and storage at -80 °C.

2.13 Lentiviral production

All lentiviral procedures were performed under Biosafety Level 2, and contaminated materials, including cells, were kept segregated from other reagents and cultures. Contaminated plasticware was deactivated by incubation in 1:10 bleach over-night, before disposal in biohazard bins for autoclaving.

2.13.1 Transfection of HEK 293T cells

As adapted from [264], HEK 293T cells ($7.6x10^6$) were seeded into 150 cm^2 flasks (*Corning*; 3291) and incubated overnight in 17 mL cIMDM, which was replaced with 19 mL cIMDM 2 hours before transfection. A mixture of pMD2.VSVG (7.6 µg), CMVΔR8.74 (13.8 µg), pRSV-REV (5.3 µg), pADvantage (12.7 µg) and transfer vector (25.5 µg) was topped up to 955.4 mL with 0.1X TE, mixed with 106.2 µL of 2.5 M $CaCl_2$ (*Sigma-Aldrich*), and 1.62 mL of 2X HBS was added drop-wise with rigorous mixing. The calcium-phosphate solution was added to the cells, incubated for 12–14 hours at 37 ˚C and replaced with 14 mL cIMDM containing 1 mM sodium butyrate (*Sigma-Aldrich*). After an additional 30 hours, the virus-containing medium was collected (as the 30-hour post-transfection harvest) and replaced with fresh 14 mL cIMDM containing 1 mM sodium butyrate. 18 hours later, the virus-containing medium was collected for a 48 hour post-transfection virus harvest, **Figure 2.8**.

0.1X TE	10X TE (10 mM Tris, 1 mM EDTA) diluted 1:10 with dH2O, filtered through *0.22*-µm filter
0.1X TE	0.03X TE:dH_2O (2:1, v/v)
2X HBS	281 mM NaCl 100 mM HEPES 1.5 mM Na_2HPO_4 pH 7.10–7.14

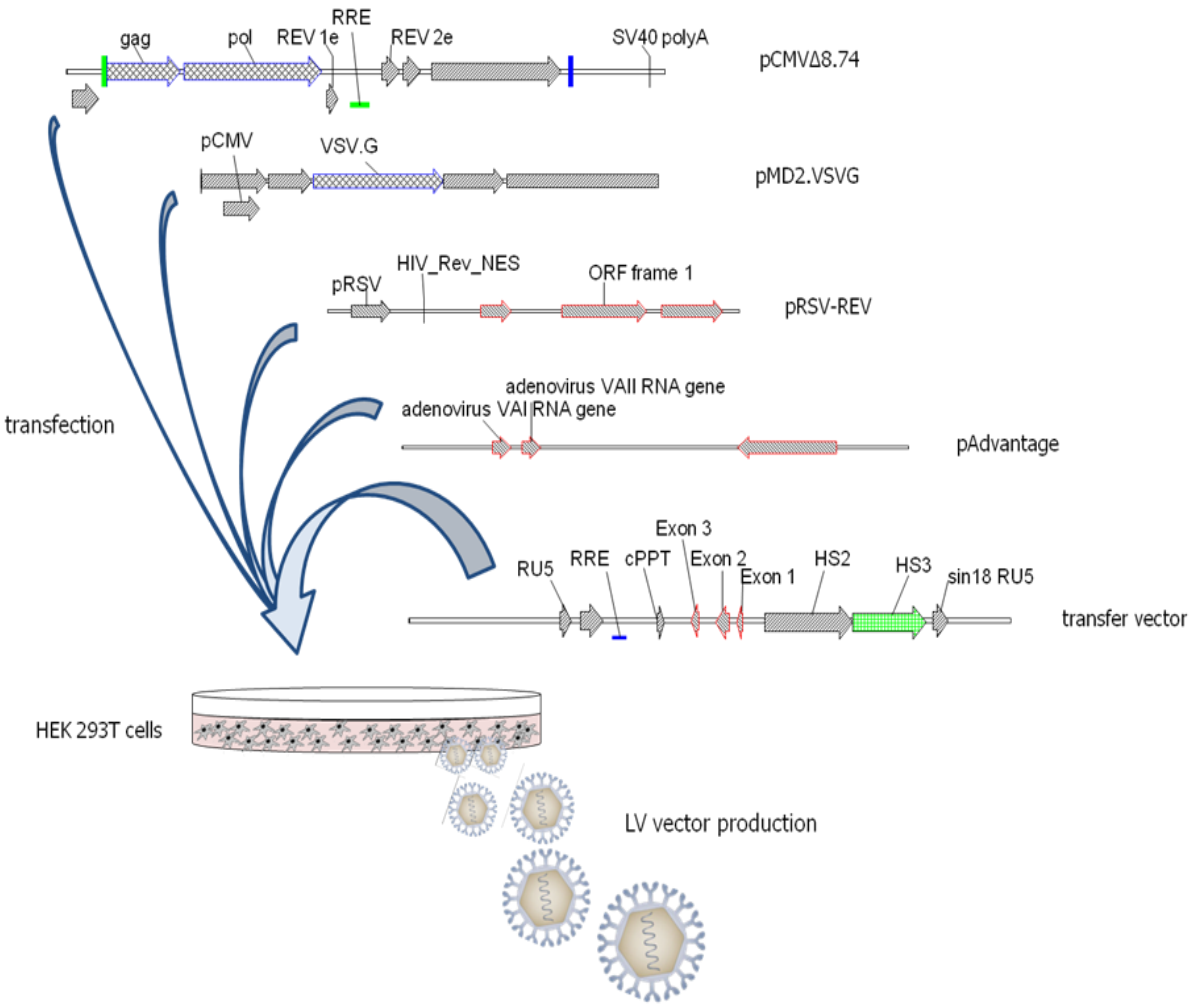


Figure 2.8. Schematic of the generation of lentiviral vectors (LVs). LV production uses a split-genome packaging system, and the relevant portions of the five plasmids co-transfected into HEK 293T cells are shown. **pCMVΔ8.74**: 2nd-generation Tat-defective packaging construct encoding functional Gag, Pol, and Rev proteins from the hCMV promoter and the SV40 polyA site. **pMD2.VSVG**: envelope plasmid encoding the heterologous vesicular stomatitis virus G glycoprotein (VSV-G) for LV pseudotyping. **pRSV-REV**: 3rd generation construct encoding the rev cDNA in order to enhance vector yield. **pAdvantage**: facultative plasmid encoding the adenovirus-associated VAI and VAII RNA genes, which enhances transient protein expression in a variety of cell types by increasing translation initiation [265]. **Transfer vector**: encodes the β-globin cDNA; the 5' LTR was modified to include the RSV promoter. CMV: cytomegalovirus; VSV-G: vesicular stomatitis virus G glycoprotein; RSV: Rous sarcoma virus; RRE: REV response element; SV40: Simian vacuolating virus 40; ORF: open reading frame; NES: nuclear export sequence; cPPT: central polypurine tract; HS: hypersensitive site; IVS: intervening region.

2.13.2 Virus concentration

Virus-containing supernatant collected at 30 hours and 48 hours post-transfection was centrifuged at 200 g for 5 minutes to settle cellular debris and filtered through a Millex-HV 0.45-*μ*m pore size PVDF membrane filter (*Millipore, Darmstadt, Germany*). Two methods of concentrating LVs particles were explored:

2.13.2.1 Ultracentrifugation

Virus was concentrated X350 by ultracentrifugation as adapted from [266] using a Beckman® Coulter Optima^{TM} L-100XP ultracentrifuge and a Beckman® SW-41 Ti rotor.

LV-containing supernatant (8 mL) was transferred into 14 x 89 mm Polyallomer centrifuge tubes (*Beckman, CA, USA*), weighed and volume adjusted to below 0.1 g difference by addition of PBS, followed by centrifugation at 50,000 g for 2 hours at 21 °C. The supernatant was decanted and tubes were positioned upside-down on absorbent paper for 10 minutes to remove residual liquid. The viral pellet was re-suspended in 1/350 of the starting volume with sterile PBS (without Ca^{2+}/Mg^{2+}), and incubated on ice at 4 °C overnight. The viral pellet was re-suspended by gentle pipetting, and viral suspensions of the same vector were combined in a single tube, aliquoted and stored at -80 °C.

2.13.2.2 High-speed centrifugation

As adapted from [267], LVs were concentrated X350 at a reduced centrifugal speed of 20,000 g at 4 °C for 2 hours on a Sigma 4K15 centrifuge. The virus-containing supernatant was transferred to conical-bottom, 50-mL tubes (*Greiner Bio-One*) fitting into a fixed-angle Sigma 12169-H rotor, and processed as described in **section 2.13.2.1**.

2.13.3 Virus titration

Vector titer (TU/mL) was determined by a limiting dilution assay on HEL cells in a 24-well plate (*Corning costar;* 3526). Briefly, 10 µL vector supernatant, which was thawed on ice, was added to 990 µL cRPMI for 1:100 dilution, and was further diluted by transferring 100 µL to 900 µL cRPMI in subsequent wells until reaching 1:10^7. Each dilution was supplemented with 8 µg/mL polybrene (*Millipore*). HEL cells were added last at a concentration of 2x10^5 cells/mL per well (**Figure 2.9**). Vector supernatant remained on cells for 24 hours, after which it was replaced with 2 mL fresh medium.

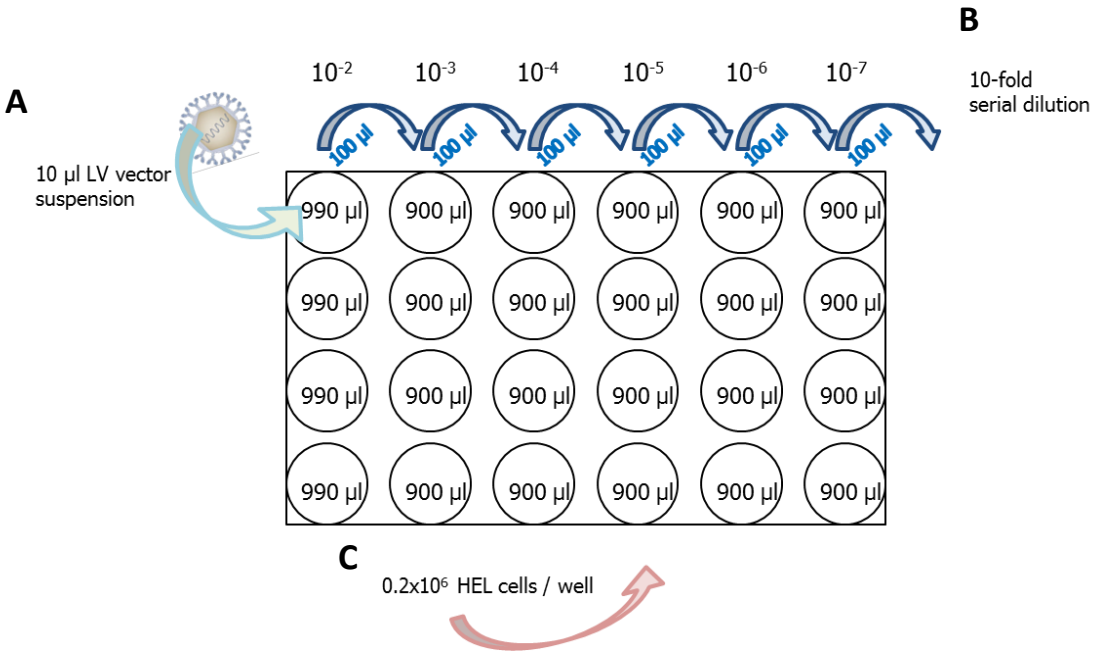


Figure 2.9. Schematic of virus limiting-dilution assay for titre determination.

2.13.4 Virus quantification techniques

2.13.4.1 Assessing viral titre by flow cytometry

For vectors containing a fluorescent reporter gene, LV quantification was performed by flow cytometry using the following formula:

$$\text{Titre (TU/mL)} = \frac{\text{initial no. of cells x (\% of GFP}^{+}\text{ cells / 100)}}{\text{LV dilution x volume of inoculum in mL}}$$

At 72 hours post-transduction, cells transduced with a PGK-GFP-expressing vector were collected, washed in PBS and suspended in 1 mL 2% FBS in PBS buffer for analysis. The viable population was gated on a plot of side scatter (SSC) vs forward scatter (FSC) for an untransduced control sample, and was further sorted based on its fluorescent properties on a histogram for channel FL1 (GFP detection), allowing 5% false-positives for GFP. All dilutions of transduced samples were analyzed for GFP expression levels, showing an increased fluorescence in the histogram of FL1 compared to control, and GFP-positive-cell counts were obtained by subtracting the negative reference counts of the control. Accurate titers can be determined if fluorescent cells receive no more than one integrant per cell. Thus, only dilutions producing 5%–10% GFP positive counts were applied in the mathematical formula above, since <5% counts will overestimate titers by including cells close to the background of false-positives, whereas, >10% counts will underestimate titers with statistically more cells having multiple integrants but not being recognised as such [268].

2.13.4.2 Assessing viral titre by qPCR

Transduced cells were maintained in culture for three weeks to eliminate non-integrated viral particles, followed by analysis for copies of stably integrated LVs in the host chromatin by qPCR using TaqMan 5' nuclease assay, **Figure 2.10**. PCRs were performed on a 7900HT Fast Real Time PCR System (*Applied Biosystems*) in parallel with known standards.

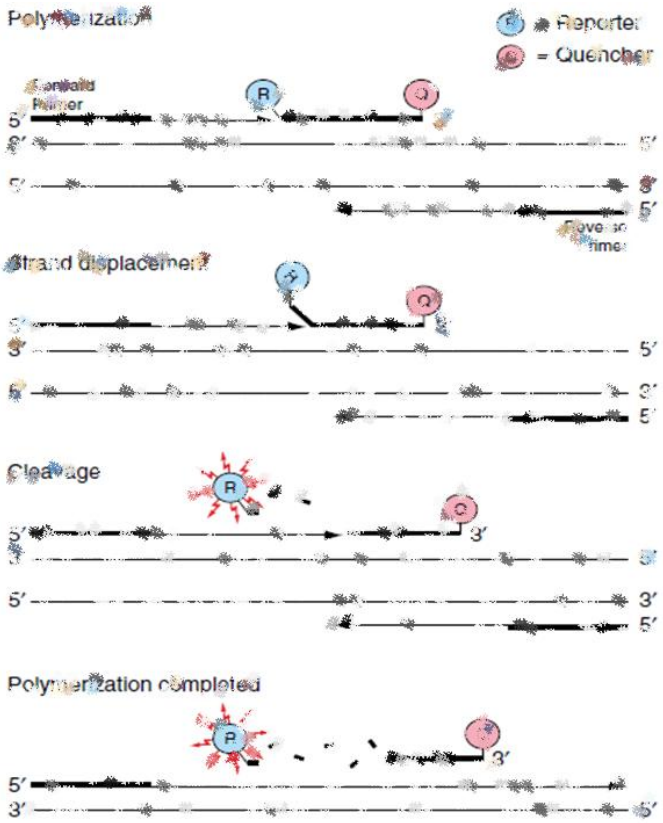


Figure 2.10. PCR amplification using TaqMan 5' nuclease assay. The assay consists of fluorogenic probes and a set of primers. TaqMan probes are covalently bound by two fluorescence dyes; a 5' reporter dye, either FAM™ (Em_{max} at 510 nm) or VIC® (Em_{max} at 554 nm), and a 3' quencher dye TAMRA (Em_{max} at 580 nm). When the probe is intact, the quencher is placed in close proximity to the reporter, suppressing its fluorescent signal. During PCR amplification, Taq polymerase cleaves the hybridized probe via its 5'-exonuclease activity, releasing the reporter dye and increasing its fluorescence intensity. With each successive amplification cycle, fluorescence intensity increases proportionally to product accumulation [269, 270]. Reproduced from [271].

The viral vector copy number (VCN) in the cell gDNA was determined from a plasmid standard curve of known qPCR copy number by comparing sample C_T values to standard equivalents. The plasmid MA821-T87Q shares common sequences for PCR detection to other lentiviral constructs and was thus chosen as a control template for VCN calculation. The mass of the plasmid was calculated based on *ABI* guidelines [272] and standards were constructed with a 5-fold dilution series from concentrated plasmid stocks using the formula $C_i \times V_i = C_f \times V_f$. Each standard was spiked with gDNA from untransduced cells in order to control for any inhibitory gDNA effects on assay performance and provide a gDNA template with similar complexity for all standard-curve, control and test samples.

Plasmid copy numbers were calculated based on the following formula

$$n = \frac{m \times N_A}{M}$$

$$m\,(g) = n\,(bp) \times 1.096e\text{-}21\,(g/bp)$$

where *n* is the DNA size (bp), *m* is the mass of the DNA, N_A is Avogardo's number (6.023e23 bp/mol), and *M* is the average molecular weight of a base pair (660 g/mol).

gDNA extracted as described in **section 2.1.1.1** was diluted in 50 ng/µL working stocks of which 2 µL were used as a template for the qPCR reaction. Therefore, the genome number present in 100 ng gDNA was calculated and further used to determine the corresponding plasmid mass needed to achieve a desired VCN, as shown below:

$$m\,(g / genome) = 3.0e{+}09\,(bp) \times 1.096e\text{-}21\,(g/bp)$$

* haploid human genome size: 3.0e+09 base pairs

$$m\,(ng / genome) = 3.29e\text{-}12\,(g) \times 10^9$$

$$\#\,genomes\ (in\ 100\ ng\ gDNA) = \frac{100\,(ng)}{3.29e\text{-}03\,(ng)}$$

$$\text{\# plasmid} = \text{desired plasmid copy number} \times \text{\# genome (in 100 ng gDNA)}$$

$$\text{plasmid mass (g)} = 9777\ \text{(bp)} \times 1.096\text{e-}21\ \text{(g/bp)}$$

$$= 1.07\text{e-}17\ \text{(g)}$$

* MA821Q plasmid size: 9777 bp

$$\text{plasmid mass (g) for known VCN} = \text{\# plasmid} \times 1.07\text{e-}17\ \text{(g)}$$

The resulting qPCR copy number was normalized to the number of genomes represented in the test gDNA sample, the latter being quantified by the quantity of endogenous PCBP2 molecules detected by qPCR in the same gDNA sample. A gDNA standard curve was generated from 5-fold serial dilutions of sample gDNA of known PCBP2 quantity at concentrations of 50 ng/µL, 10 ng/µL, 2 ng/µL, and 0.4 ng/µL.

Simplex qPCR amplification for viral DNA and gDNA was performed with 1X TaqMan buffer (*Invitrogen*^{TM}; 4318157), 900 nM forward and reverse primer, 250 nM fluorescent probe, and 100 ng gDNA in a final reaction volume of 25 µL with the following conditions: 10 minutes 95 °C, (15 seconds 95 °C, 1 minute 60 °C) x 40 cycles. In each PCR run, standards and negative controls were run in triplicate and experimental samples were run in duplicate.

Primer and probe sequences (*Metabion*) are as shown below:

Name	**Sequence 5' to 3'**
Lenti FW	TGAAAGCGAAAGGGAAACCA
Lenti RV	TTGCCGTGCGCGCTTCAG
Lenti probe	VIC-AGCTCTCTCGACGCAGGACTCGGC-TAMRA
PCBP2_FW	CTGCATAATCGGGCGTCAAG
PCBP2_RV	GCAGCAGATCCAGTGATGGTAACCT
PCBP2_probe	FAM-CGCCAAAATCAATGAGATCCGTCAGATGTCT-TAMRA

Quantification was carried out by 7900 Fast System SDS 2.4 software (*Applied Biosystems*), and the viral copy number per genome was calculated using *Microsoft® Excel* according to the equation

$$\text{Vector titre (TU/mL)} = \frac{\text{Lenti quantity x 100 ng}}{\text{PCBP2 quantity}} \text{ x dilution factor x cell number}$$

For accurate titre values, only titres from transductions with dilution factors producing 5–15-fold change in qPCR copy number, assuming 10-fold dilution standards, were calculated and averaged to yield a mean total copy number value.

2.14 Statistical analysis

For each data set (group of raw values) a normality test was conducted (see below) to assess data distribution. If shown to follow a Gaussian function, a parametric test (t-test, one-way ANOVA and two-way ANOVA) was used under the assumption that group values were normally distributed. Statistical tests give the P value, which determines the statistical significance in a hypothesis test. If the test is significant (p<0.05), the null hypothesis (the possibility that there is no difference between groups) is rejected. The t-test and ANOVA calculate the t-value and F-ratio, respectively, which are used to compute the P-value. Outliers from a set of data were removed automatically before statistical analyses (all *Microsoft Excel*), choosing a deviation of two standard deviations from the sample mean as a cut-off value.

The Prism 5.0 software (*GraphPad Software Inc., CA, USA*) offers the choice of three normality tests [273]; D'agostino-Pearson (n≥8), Shapiro-Wilk (n≥7), and Kolmogorov-Smirnov (n≥5). If the test was significant (p<0.05), the null hypothesis of normal sample distribution was rejected. Small data groups increase the risk of having lower power to detect modest deviations from Gaussian distribution, and might instead be assessed for skewness and kurtosis to determine normality [274, 275]. If tests for normality were negative and logarithmic data transformation did not result in (approximate) normality of the data [275], then the non-parametric Kruskal-Wallis test was applied to group comparison.

3.0 Results

3.1 Lentiviral vector construction

3.1.0 Aims

- To design simple shRNA and miRNA-adapted shRNA hairpins.
- To produce modified GLOBE plasmids carrying shRNA-expressing constructs in the βIVS-II region of the *HBB* transgene cassette.
- To produce pLKO.1-based plasmids encoding shRNAs under the RNA-pol-III U6 promoter.

3.1.1 Introduction

Lentiviruses are members of the *Retroviridae* family, species which use viral reverse transcriptase (RT) to produce a double stranded DNA (dsDNA) copy from their RNA genome and integrase (IN) to stably integrate the newly synthesised dsDNA into the host genome [276, 277]. Lentiviruses were proposed in the mid-1990s as gene therapy (GT) vectors mainly because of their ability to produce sustained gene expression through vector integration, deliver relatively large and complex transgene cassettes, target different cell types through pseudotyping, infect both dividing and non-dividing cells, and abolish production of full-length vector RNA in transduced cells through a self-inactivating (SIN) design [264, 278]. One species of lentivirus, the human immunodeficiency virus (HIV-1), has been proposed as an attractive gene delivery vehicle because, unlike MoLoney murine leukemia (MLV)-based vectors, it does not show an integration preference around promoters of actively transcribed genes but throughout the transcription units [171, 279].

HIV-1 has a 9 kb single-stranded positive-sense RNA genome that encodes nine viral proteins. The viral genome is located within a nuclear core surrounded by a nucleocapsid and viral envelope core. The three largest open-reading frames encode its three major structural proteins Gag, Pol and Env. The *gag* gene encodes the capsid proteins. The *pol* gene encodes the viral enzymes reverse transcriptase, integrase and protease, which are required for viral replication.

These three enzymes are expressed within the context of a Gag-Pol fusion protein. The protease activity is required to cleave the Gag-Pol precursor during virion maturation and further digest Pol to separate the other enzymatic activities. The *env* gene encodes the viral surface glycoprotein gp160, which is required for infectivity. Also, the viral genome encodes the regulatory proteins Tat and Rev, which activate viral transcription and control the nuclear export of viral transcripts, respectively. The *rev* gene promotes its activity through interaction with the *rev* response element (RRE). The HIV-1 specifically contains four additional accessory genes namely *vif*, *vpr*, *vpu*¸ and *nef*, which are critical for *in vivo* replication and pathogenesis. A detailed account on the function of these genes can be found at http://hivinsite.ucsf.edu/InSite?page=kb-02-01-02#S2.1.2X.

The viral genome is flanked by long terminal repeats (LTRs) consisting of *cis* elements required for gene expression, reverse transcription and integration into the host chromosome. Other important elements include the psi (Ψ)-sequence for RNA encapsidation and formation of new virions, and the central polypurine tract (cPPT) as the site of cDNA synthesis initiation for the plus DNA strand during reverse transcription. The double-stranded DNA is able to integrate in the genome of this host and in this form is referred to as the 'provirus' (**Figure 3.1.1**) [277, 280, 281].

LVs used in GT are developed following SIN modifications to prevent vector DNA mobilisation and recombination with wild-type virus. Construct modifications to improve safety include deletion of genes critical for virulence and pathogenicity, partition of genes into separate packaging plasmids, and deletion of enhancer/promoter sequences in the U3 region of the 3' LTR [280, 282].

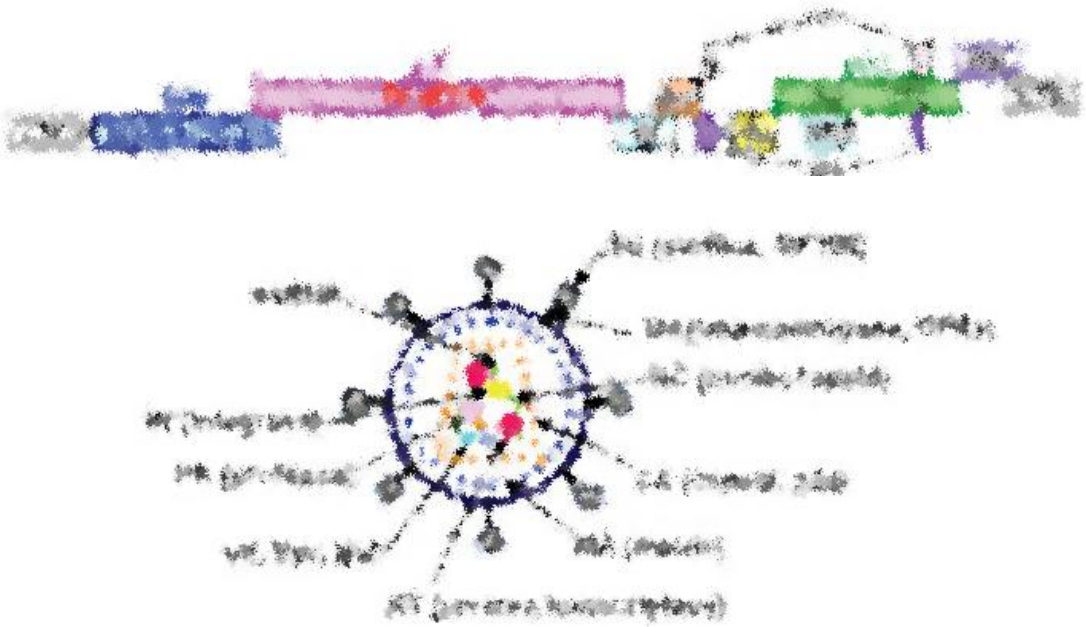


Figure 3.1.1. Schematic of the HIV-1 wild-type viral genome and structure. After reverse transcription from RNA to dsDNA the proviral genome is flanked by two LTR sequences at each end. The viral genome encodes three structural (*gag, pol, env*), two regulatory (*tat, rev*), and five accessory (*vif*, *vpr*, *vpu*¸ *nef*) genes. The Gag precursor is processed to matrix (MA; p17), capsid (CA; p24), nucleocapsid (NC; p7), and p6 proteins. The Gag-Pol precursor is processed to reverse transcriptase (RT), integrase (IN), and protease (PR) enzymes. The Env glycoprotein precursor gp160 is processed to the external glycoprotein gp120 (SU domain) and the transmembrane glycoprotein gp41 (TM domain). The rev responsive element (REV) is also shown. Reproduced from [281].

3.1.2 Strategy design of RNAi-based therapeutics

The basic strategy consists of modification of GLOBE, an LV expressing *HBB* under the late-erythroid *HBB* promoter and βLCR elements HS2 and HS3 [156], to include shRNAs in the intervening IVS-II region, which upon expression would target negative regulators of *HBG* expression and re-activate HbF production (**Figure 3.1.2a**). The shRNA cloning strategy follows the work of Sadelain and colleagues [224], such that our shRNA insertion site was designed to be located in a very similar position to the one published near the branch-point domain of IVS-II (**Figure 1.14**). As previously described (**section 1.6.1**), the endogenous RNAi pathway includes intronic miRNAs. In the nucleus, the miRNA is co-expressed with its encoding gene from the same RNA-pol-II promoter. The splicing machinery catalyses the excision of introns from the

pre-mRNA transcript and facilitates the fusion of exons to produce mature mRNAs for protein synthesis. The excised introns that contain lariat-encoded miRNAs are digested into monoribonucleotides for transcriptional recycling. However, it has been noted that about 10%–30% of the intron remains intact, serving as a source for miRNAs, which can be processed by the RNAi machinery to produce functional siRNAs [202]. Therefore the design of artificial introns carrying hairpin-like RNAi triggers appears to be an attractive therapeutic strategy.

HBB is composed of three exons and two introns. Introns contain elements critical for gene regulation and both βIVS-I and βIVS-II are needed to achieve maximum production of the mature transcript with the second intron (βIVS-II) being particularly important in this regard, implicated in promoting mRNA 3' end formation (cleavage, polyadenylation) [283, 284]. Studies narrowed this effect to within sequences of the terminal 60nt of βIVS-II, in particular the polypyrimidine tract between the lariat branch-point and the 3' splice acceptor site with the U2AF65 factor binding to the polypyrimidine tract playing a crucial role in this process emphasizing the importance of the number and position of constituting purine residues [284, 285].

Additional studies identified regions within the βIVS-II implicated in low retroviral vector production through an early termination of viral transcription. Deleting this portion of βIVS-II, while retaining intron-mediated elements essential for *HBB* expression, resulted in higher levels of full-length viral RNA and accumulation of vector particles at higher titers [286, 287]. To this end, the *HBB* cassette inserted in the GLOBE vector carries a deletion (562bp [174]) in the βIVS-II, while the full-length βIVS-I is retained.

Drawing on from this information, we designed an shRNA-expressing cassette and inserted this in βIVS-II assuming that a deleted region would accept the extra sequence without any adverse effects on virus production and/or transgene expression, which could result from perturbation of the natural sequence and thus of elements involved in transcription and/or RNA splicing. The *HBB* transgene expression cassette was inserted in reverse orientation in the GLOBE LV to prevent removal of introns by splicing during virus production as well as to maintain spatial and temporal expression under the *HBB* promoter/βLCR elements. The choice of insertion sites

within βIVS-II was limited by availability of restriction enzyme sites in what could have been a suitable position in the region. Using the Vector NTI software (*Life Technologies*), we searched the plasmid sequence map for all restriction enzyme sites available within the GLOBE βIVS-II. The query revealed a unique *Bsr*GI as a suitable site, the only one in a region unrecognized by splicing components. To ensure siRNA potency via comparative analysis we constructed an artificial restriction enzyme site in βIVS-II. We screened the intron for known mutations using the UCSC Genome Browser (http://genome.ucsc.edu/) and selected a region close to the branch point, which we presumed suitable for modification. The sequence at this position was modified to a *Bsi*WI recognition site, which would produce compatible cohesive ends (5'-GTAC) to *Bsr*GI, and would therefore facilitate easier insertion of the shRNA-expressing cassette.

The shRNA cassette was thus flanked by *Bsr*GI sites for ligation at the βIVS-II *Bsr*GI site. However, unforeseen cloning difficulties significantly delayed vector construction and subsequently *in vitro* characterization. Therefore once sequence-confirmed vectors were at hand, we proceeded to their functional testing. In parallel, we initiated efforts to transfer the shRNA-expressing cassette into the engineered *Bsi*WI site within βIVS-II, which was later abandoned mainly because of time-consuming technical cloning problems at the expense of establishing a functional *in vitro* culture system and partly because of concerns raised over the functionality of the *Bsi*WI position. A survey of the literature revealed that *Bsi*WI is embedded within a region recognized by the splicing factors U2AF and SF1/BBP, as well as the small nuclear ribonucleoprotein (snRNP) U2 [285, 288, 289]) during splicing and processing. The spliceosome catalyzes intron removal from pre-mRNAs by sequential assembly of U1-U5 snRNPs and numerous other splicing factors [290, 291] and by their ordered interaction with conserved sequences within the intron causing structural rearrangement during catalytic steps of splicing (for information on intronic splicing sites recognized by spliceosomal components see [285, 288, 289, 292-299]). The splicing recognition sites in the intron lie within conserved sequences of splice sites (3' splice acceptor and 5' splice donor) and a branch-point domain followed by a polypyrimidine tract (**Figure 3.1.2b**). Particularly, a snRNP interaction with the branch-point domain facilitates formation of the pre-spliceosome, which catalyses the formation of an intermediate lariat-like structure essential for subsequent intron splicing. In our design, the

*Bsr*GI site was positioned approximately 135bp upstream of the branch-site and with appreciable distance to any of the other conserved sequences. However, *Bsi*WI was created at approximately 19bp downstream of the branch-site domain with unintentional modification of the last polypyrimidine tract base from a pyrimidine (Cytosine; C) to a purine (guanine; G). The creation of the *Bsi*WI site required modification of another base in a region, which appears to foster binding of spliceosomal snRNP complexes essential for a branch-site interaction. This information raises the possibility of resulting in reduced splicing efficiency, although not conclusively. At that stage, we did not have the time to reverse modifications by restoring the initial βIVS-II sequence nor to re-create insertion sites at possibly better positions within the region. To this end, we proceeded with evaluation of the *Bsr*GI site as a means of providing a proof-of-principle for our shRNA-based therapeutic strategy.

Figure 3.1.2. A) Schematic representation of the shRNA-expressing MA821Q LV in its proviral form. **B)** Sequence map of the *HBB*^{T87Q} transgene expressed from the MA821Q LV. The DNA sequence of the three exons, the two intervening sequence regions (introns) and the cPPT element are shown. Features (restriction enzyme sites, snRNP binding sites, conserved intronic regions and cleavage/ poly(A)-addition elements) are color-coded as text-based graphics. cPPT: central polypurine tract; IVS: intervening region; SD: splice donor site; SA: splice acceptor site, PTB: pyrimidine tract binding protein; USE: U-rich upstream sequence element; DSE: U-/GU-rich downstream sequence element (USE and DSE regions bind hnRNP proteins, facilitating 3' end formation [292]).

3.1.3 Construction of shRNA-based vectors

A range of candidate vectors were constructed based on the *HBB*-expressing LV MA821 (GLOBE) [156], and modified to hold the innocuous anti-sickling *HBB* variant T87Q for reliable follow-up analyses of LV-derived expression [236], (data not shown). The cloning strategy for expressing shRNA constructs from the GLOBE vector is illustrated in **Figure 3.1.3**.

In the following, vector constructs will be referred to in short-hand, as follows:

MA821-T87Q	MA821Q
MA821-T87Q-SF	MA821Q_SF
MA821-T87Q-LF	MA821Q_LF
MA821-T87Q-BCL11A(shRNA449)	MA821Qsh_B9
MA821-T87Q-BCL11A(miRNA449)	MA821QmiR_B9
pLKO-shRNA BCL11A(449)	pLKOsh_B9
MA821-T87Q-BCL11A(shRNA451)	MA821Qsh_B1
MA821-T87Q-BCL11A(miRNA451)	MA821QmiR_B1
pLKO-shRNA BCL11A(451)	pLKOsh_B1
MA821-T87Q-shRNA GFP	MA821Qsh_G
pLKO-shRNA GFP	pLKOsh_G
MA821-T87Q-shRNA BCL11A(451)-shRNA GFP	MA821Qsh_B1_G
MA821-T87Q-shRNA BCL11A(451)-shRNA BCL11A(4449)	MA821Qsh_B1_B9
MA821-T87Q-miRNA Scramble	MA821QmiR_Scr
pLKO-shRNA Scramble	pLKOsh_Scr

The HBB-βLCR cassette was transferred from MA821Q as a 2587bp *Cla*I/*Not*I fragment into the pBluescript (*Stratagene,* La Jolla, CA) plasmid, generating pBS_HBBQ.

Synthetic DNA fragments designed to encode multiple shRNAs were inserted at the junction point of the internal βIVS-II deletion of the MA821Q plasmid, 173bp upstream of the exon 3 start site, with each shRNA flanked by a unique dual restriction site. While the design of the full shRNA cassette (**Figure 3.1.4**) would facilitate subsequent production of plasmids carrying

individual or combinations of shRNAs, the presence of six shRNA-encoding DNA elements in this large fragment (LF) delayed its synthesis, cloning and confirmation by the manufacturer. A reduced construct bearing only BCL11A and GFP shRNAs (small fragment, SF) was thus synthesized to expedite initial analyses (**Figure 3.1.5**).

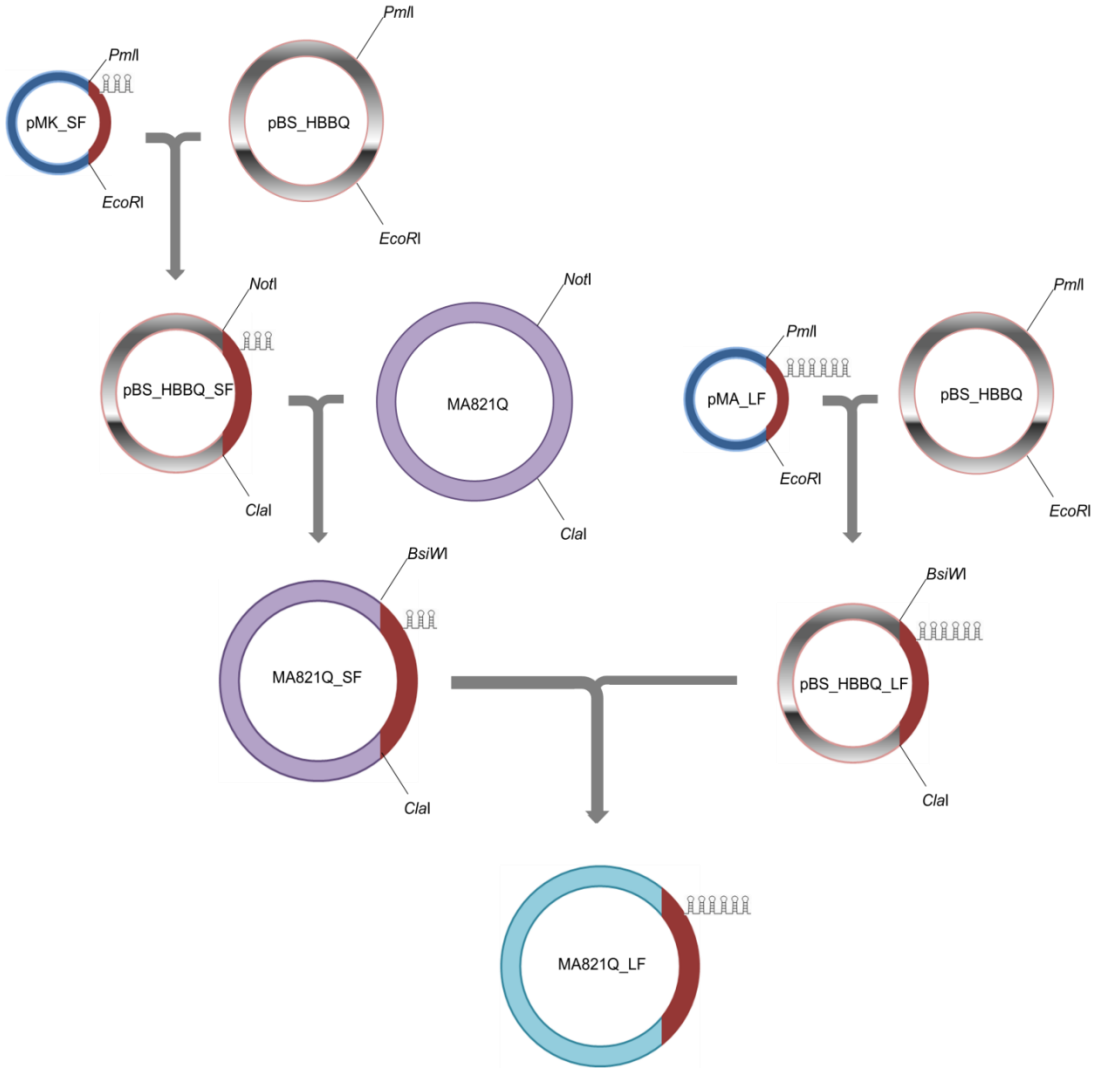


Figure 3.1.3. Cloning strategy for MA821Q_SF and MA821Q_LF production. The diagram illustrates the cloning of a three-shRNA-encoding fragment (SF) and a larger six-shRNA-encoding fragment (LF) into the MA821Q vector. The shRNA cassettes were initially cloned into pBluescript and then re-cloned into the MA821Q. The pBluescript plasmid contains a wider choice of unique restriction sites and thus allows design of a smaller synthetic fragment for sub-cloning, compared to the full-length MA821Q plasmid. Cloning of the LF into the MA821Q plasmid was facilitated by additional restriction sites available in the MA821Q_SF plasmid.

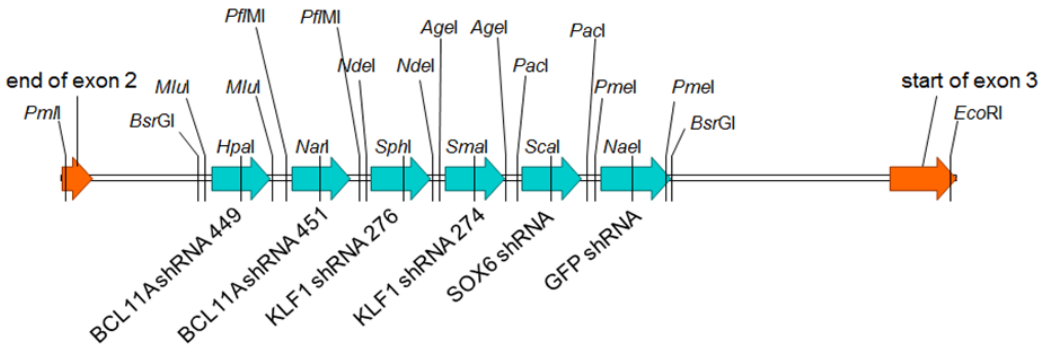


Figure 3.1.4. Schematic of LF. The synthetic fragment contains six shRNA sequences specific for GFP, human BCL11A (two validated shRNAs), KLF1 (2 validated shRNAs) and SOX6. It covers *HBB* IVS-II from the *Pml*I site at the 3' end of exon 2 to the *Eco*RI site at the 5' end of exon 3.

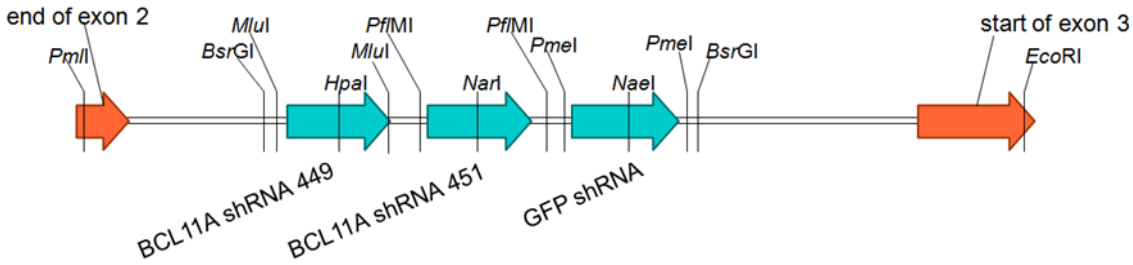


Figure 3.1.5. Schematic of SF. The synthetic fragment contains three shRNA sequences specific for GFP and human BCL11A (two validated shRNAs). It covers *HBB* IVS-II from the *Pml*I site at the 3' end of exon 2 to the *Eco*RI site at the 5' end of exon 3.

The siRNA sequences on target genes were chosen empirically from published work that showed efficient knockdown by shRNA expression constructs, given that, as yet, the molecular mechanisms underlying RNAi are not fully elucidated to enable rational design of siRNA hairpins. Publically available state-of-the-art tools, which were proposed to accurately predict siRNA potency have instead shown variable performance in predicting siRNA activity with algorithm sensitivity and specificity affected by the set of features implemented in the design process [300, 301]. Candidate shRNAs were designed to incorporate the validated siRNA duplex sequences with a spacer sequence (the loop) encoding for unique restriction enzyme sites according to the shRNA design guidelines of the RNAi consortium library [259]. The

efficacy of shRNAs is hairpin-dependent but also gene and target-cell line-dependent. Therefore, considering a plausible variability in silencing activity among same-gene candidate siRNA, we used two test shRNAs per gene in primary screen assays in order to increase the probability of obtaining robust knockdown of target genes (**Table 3.1.1**).

Name	sense strand	loop	anti-sense strand
shRNA BCL11A(449)	CGCACAGAACACTCATGGATT	GTTAAC	AATCCATGAGTGTTCTGTGCG
shRNA BCL11A(451)	CCAGAGGATGACGATTGTTTA	GGCGCC	TAAACAATCGTCATCCTCTGG
shRNA KLF1(276)	CAGAGGATCCAGGTGTGATA	GCATGC	TATCACACCTGGATCCTCTGC
shRNA KLF1(274)	CCCGAGACTCTGGGCGCATAT	CCCGGG	ATATGCGCCCAGAGTCTCGGG
shRNA SOX6	CCAGTGAACTTCTTGGAGAAA	AGTACT	TTTCTCCAAGAAGTTCACTGG
shRNA GFP	GCAAGCTGACCCTGAAGTTC	TTCGCCGGC	GAACTTCAGGGTCAGCTTGC

Table 3.1.1. The siRNA target sequences for BCL11A(449) [63, 119], BCL11A(451), KLF1(276), KLF1(274) [142], SOX6 [63] and GFP [302]. All sequences are shown 5' → 3'.

3.1.3.1 Construction of MA821Q_SF

3.1.3.1.1 Cloning SF into the pBS_HBBQ vector

SF (585bp), encoding the two BCL11A shRNAs and the control GFP shRNA, was received as a 2843bp construct inserted into the *Sfi*I site of the pMK (kanR) vector. The pMK-SF plasmid was transformed into XL1 Blue *E. coli* cells and correct clones were determined by restriction enzyme digestion with *Sfi*I (**Figure 3.1.6)**.

In order to construct the pBS_HBB_T87Q-SF plasmid, the shRNA-encoding DNA was excised from pMK-SF as a 534bp *Pml*I/*Eco*RI fragment and inserted into the corresponding sites of pBS_HBBQ. The plasmid vector cloning sites *Pml*I and *Eco*RI were selected based on their

proximity to the shRNA cassette in the construct provided by the manufacturer (GeneArt) and on their properties as unique cutting sites within the pBS_HBBQ plasmid.

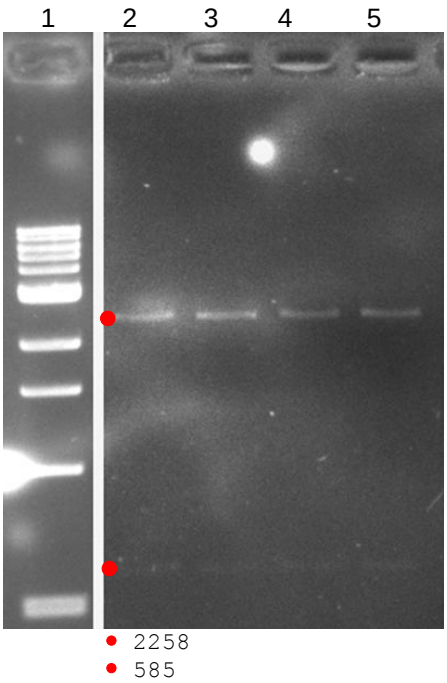


Figure 3.1.6. Agarose gel separation of *Sfi*I-digested pMK_SF clones confirming their identity. Analytical agarose gel electrophoresis of pMK_SF clones [2–5] digested with *Sfi*I. [1] 1-kb ladder. •: fragments of interest (bp).

The SF insert was ligated into the target plasmid at a molar ratio of 3:1, and the ligation reaction (5 µl) was used to transform XL1 Blue *E. coli* cells. An aliquot of 500ng plasmid DNA isolated from candidate clones was digested with *Hpa*I to confirm the presence of the insert and to determine its orientation within the plasmid. In the pBS_HBBQ_SF plasmid, there is one *Hpa*I site asymmetrically placed within the insert 392bp upstream the *Eco*RI site, with another *Hpa*I site located within the vector 636bp downstream the *Eco*RI site. Agarose gel analysis of the *Hpa*I digested cloned plasmid DNA confirmed insertion of the SF insert in the correct orientation, showing a 1028bp and a 4670bp fragment, while an insert in the reverse orientation would have resulted in 781bp and 4917bp fragments (**Figure 3.1.7**).

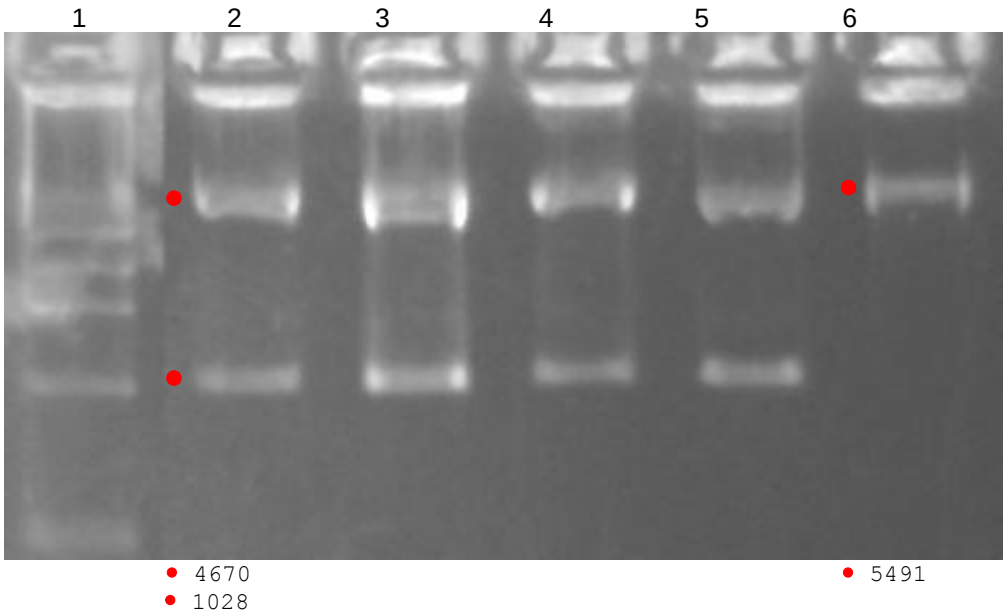


Figure 3.1.7. Agarose gel separation of *Hpa*I-digested pBS_HBBQ_SF clones. Analytical agarose gel electrophoresis of pBS_HBBQ_SF [2–5] and pBS_HBBQ [6] digested with *Hpa*I. [1] 1-kb DNA ladder. ●: fragments of interest (bp).

3.1.3.1.2 Cloning SF into MA821Q

In order to construct the MA821Q_SF plasmid, the shRNA-encoding DNA was excised from the 5698bp pBS_HBBQ_SF as a 2794bp *Not*I/*Cla*I fragment, and ligated into the same sites of the 9771bp MA821Q.

The ligation reaction was used to transform *Stratagene* XL10-Gold® Ultracompetent Cells. Candidate clones were verified by digestion with *Hpa*I to test for the presence of the insert in the correct orientation at the *Not*I/*Cla*I insertion site. Agarose gel analysis identified three of the screened MA821Q_SF clones as having the correct vector (8950 bp) and insert (1028 bp) sizes (**Figure 3.1.8**). The correctness of all three of these clones was further confirmed by digestion with *Bsr*GI (**Figure 3.1.9**).

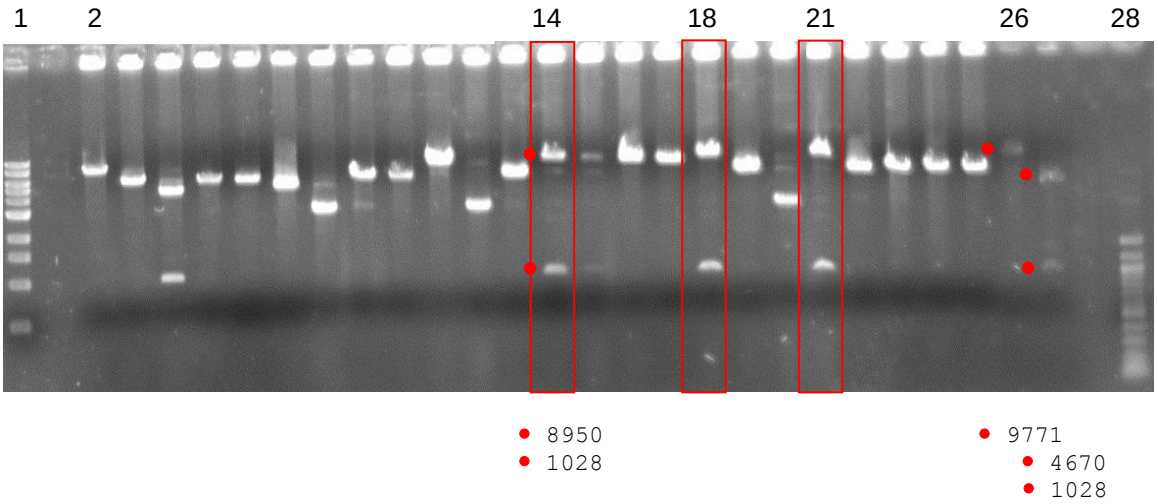


Figure 3.1.8. Agarose gel separation of *Hpa*I-digested MA821Q_SF clones. Analytical agarose gel electrophoresis of MA821Q_SF candidate clones [2–25], MA821Q [26], and pBS_HBBQ_SF [27] digested with *Hpa*I. Correct clones ([14], [18] and [21]) are boxed in red. [1] 1-kb DNA ladder. [28] 100-kb DNA ladder. •: fragments of interest (bp).

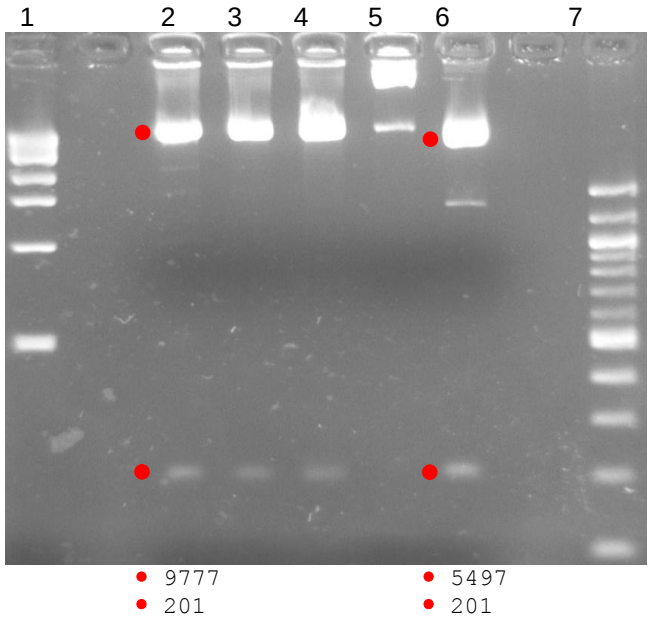


Figure 3.1.9. Agarose gel separation of *Bsr*GI-digested MA821Q_SF clones. Analytical agarose gel electrophoresis of MA821Q_SF candidate clones [2–5], and pBS_HBBQ_SF [6] digested with *Bsr*GI. [1] 1kb DNA ladder. [7] 100-kb DNA ladder. •: fragments of interest (bp).

In addition to restriction digestion analysis, the MA821Q_SF clone identity was also confirmed by sequencing the insert for *HBB* using vector primers flanking the insertion sites (**Figure 3.1.10**). Primer sequences are shown in **section 2.3**. Coverage of >80% of the *Cla*I/*Not*I fragment was achieved by using internal primers, which, nonetheless, produced a poor

sequencing trace across a 360bp region of vector backbone downstream exon 3. Complete coverage of vector backbone was not explored further given the fact that the plasmid was previously sequenced by its manufacturer.

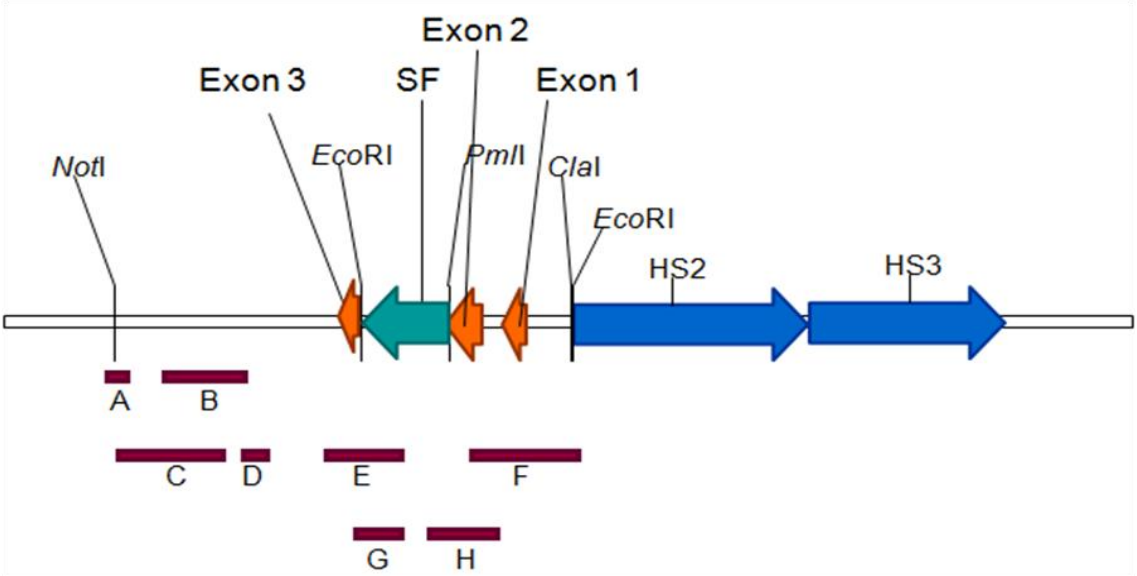


Figure 3.1.10. Sequence analysis of MA821Q_SF. Diagram of sequencing primers for *HBB* in the MA821Q_SF vector. The position of the primers around the *Pml*I/*Eco*RI and *Not*I/*Cla*I insertion sites, and the read length relative to the construct sequence is shown. A: MA821_NotI_FW; B: 3402_SEQ; C: 3099_SEQ; D: 3861_SEQ; E: ARMS 10; F: MA821_ClaI_RV; G: IVS2_shSEQ_RV; H: αPCO4_FW.

3.1.3.2 Construction of the MA821Qsh_B1_G vector

The MA821Q-B1_G vector was generated by excising the 54bp BCL11A_449-shRNA from the 9978bp MA821Q_SF by *Mlu*I digestion, vector self-religation and transformation in DH5α™ competent cells. After culture and plasmid isolation from candidate clones, DNA was digested with *Hpa*I and *Nar*I. Both enzymes cut once within the vector backbone and have additional recognition sites in the BCL11A_449-shRNA and the BCL11A_451-shRNA loops, respectively. Therefore correct clones were identified by the generation of a single 9924bp fragment after *Hpa*I digestion (as the BCL11A_449-shRNA-specific site will have been lost) and two fragments (of 7576bp and 2348bp) after *Nar*I digestion. This gave three clones that appeared to be correct

(**Figure 3.1.11**), which were confirmed by further digestion with *Pfl*MI and *Pme*I. Correctly verified clones were sequenced using the IVS2_shSEQ primer set (**section 2.3**). Standard sequencing reactions stopped within the first 10 nucleotides of the outermost hairpin sequences, thus confirming the absence of the BCL11A_449 shRNA and the integrity of the sequences flanking the deletion (data not shown).

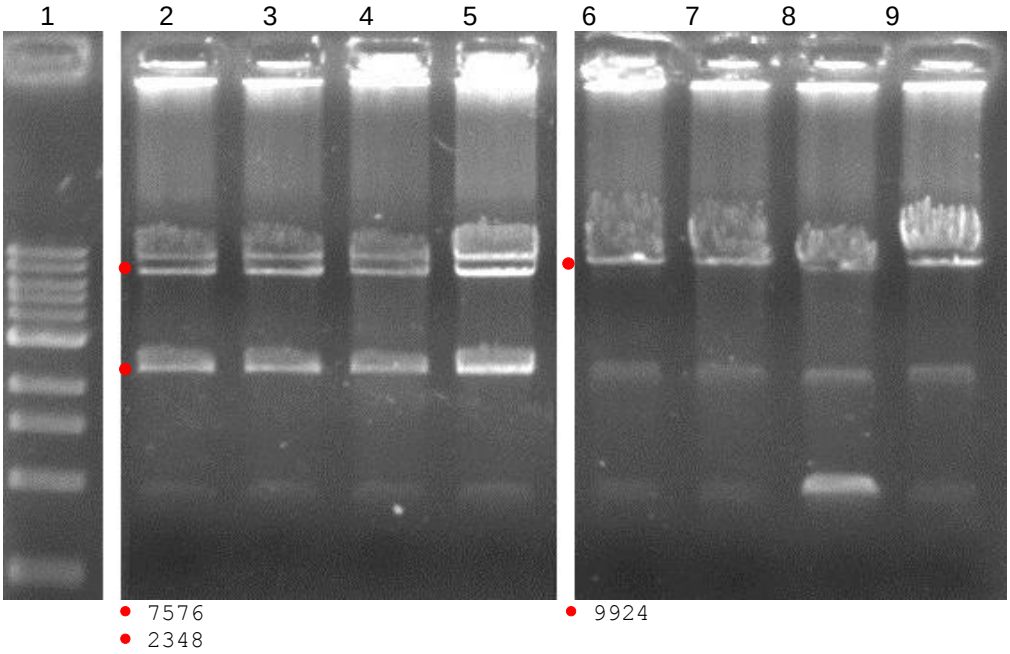


Figure 3.1.11. Agarose gel separation of MA821Qsh_B1_G digested with *Nar*I and *Hpa*I. Analytical agarose gel electrophoresis of four MA821Qsh_B1_G candidate clones digested with *Nar*I [2–5] and *Hpa*I [6–9]. [1] 1-kb DNA ladder. ●: fragments of interest (bp).

3.1.3.3 Construction of the MA821Qsh_B1 and MA821Qsh_G vector

Based on the 9924bp MA821Qsh_B1_G plasmid, MA821Qsh_B1 was generated by excision of the 57bp GFP-shRNA by *Pme*I digestion and MA821Q_G by removal of the 59bp BCL11A_451-shRNA by cutting with *Pfl*MI followed by vector self-religation and transformation in DH5α^{TM} competent cells. Plasmids from candidate clones were digested with *Nae*I and *Nar*I to confirm the presence and the absence of shRNA elements respectively. A lack of a second *Nae*I site confirmed the absence of the GFP-shRNA in MA821Qsh_B1 and a lack of a second *Nar*I site demonstrates the absence of the BCL11A_451 shRNA in MA821Qsh_G (**Figure 3.1.12**).

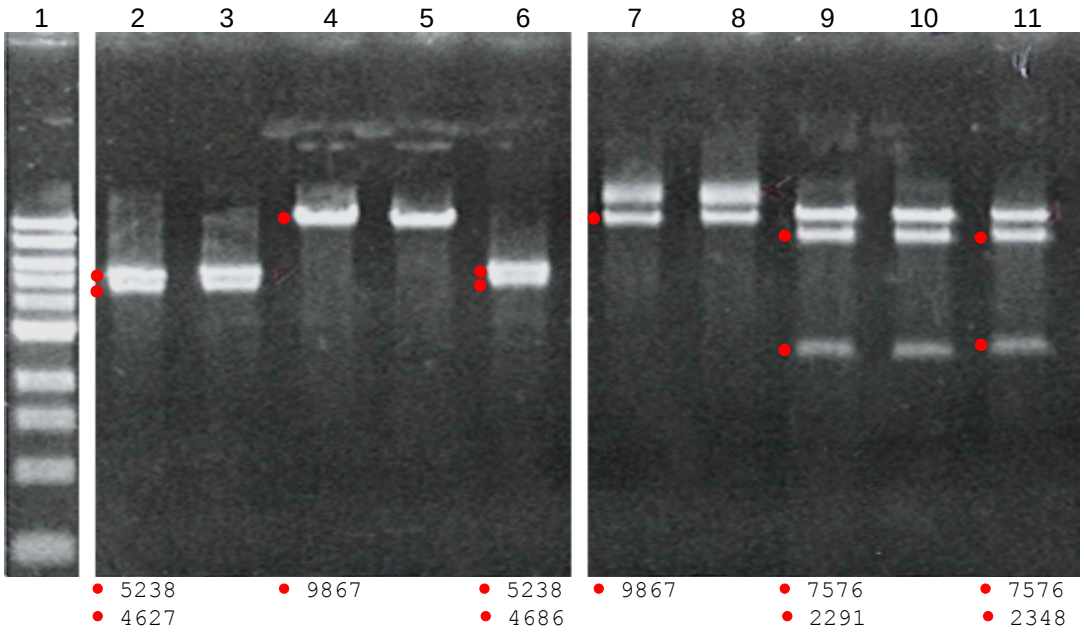


Figure 3.1.12. Agarose gel separation of MA821Qsh_G and MA821Qsh_B1 digested with *Nae*I and *Nar*I. Analytical agarose gel electrophoresis of MA821Qsh_ G, MA821Qsh_B1, and MA821Qsh_B1_G digested with *Nae*I [2–3, 4–5, 6] and *Nar*I [7–8, 9–11, 11]. [1] 1-kb DNA ladder. •: fragments of interest (bp).

3.1.3.4 Construction of the MA821Qsh_B1_B9 vector

The MA821Qsh_B1_B9 vector was generated by excision of the 57bp GFP-shRNA from the 9978bp MA821Q_SF by digestion with *Pme*I followed by plasmid vector re-ligation and transformation of DH5α^{TM} competent cells. Candidate bacterial clones were sub-cultured and assayed by digestion with *Hpa*I, *Nar*I (confirmed presence of BCL11A_449-shRNA and BCL11A_451-shRNA loops, respectively) and *Nae*I (confirmed absence of GFP-shRNA loop). Correct clones were expected to produce two fragments of 8950bp and 1028bp after *Hpa*I digestion and of 7630bp and 2348bp after *Nar*I digestion. Plasmid linearization (yielding a 9978bp product) was produced after *Nae*I digestion due to lack of a second restriction site for this enzyme due to excision of the GFP-shRNA (**Figure 3.1.13**). Positive clones were verified by sequencing using the IVS2_shSEQ primer set. In order to overcome the sequencing block at the secondary structural motif, DNA was digested across the stem loop using *Hpa*I or *Nar*I resulting in sequencing half of the template from the forward direction and the other half from the reverse direction. Sequencing reactions stopped 25bp into the BCL11A_451-shRNA loop

and covered 24bp at the end of the BCL11A_449-shRNA loop, thus confirming the integrity of the two BCL11A hairpins and the absence of the GFP-shRNA (data not shown).

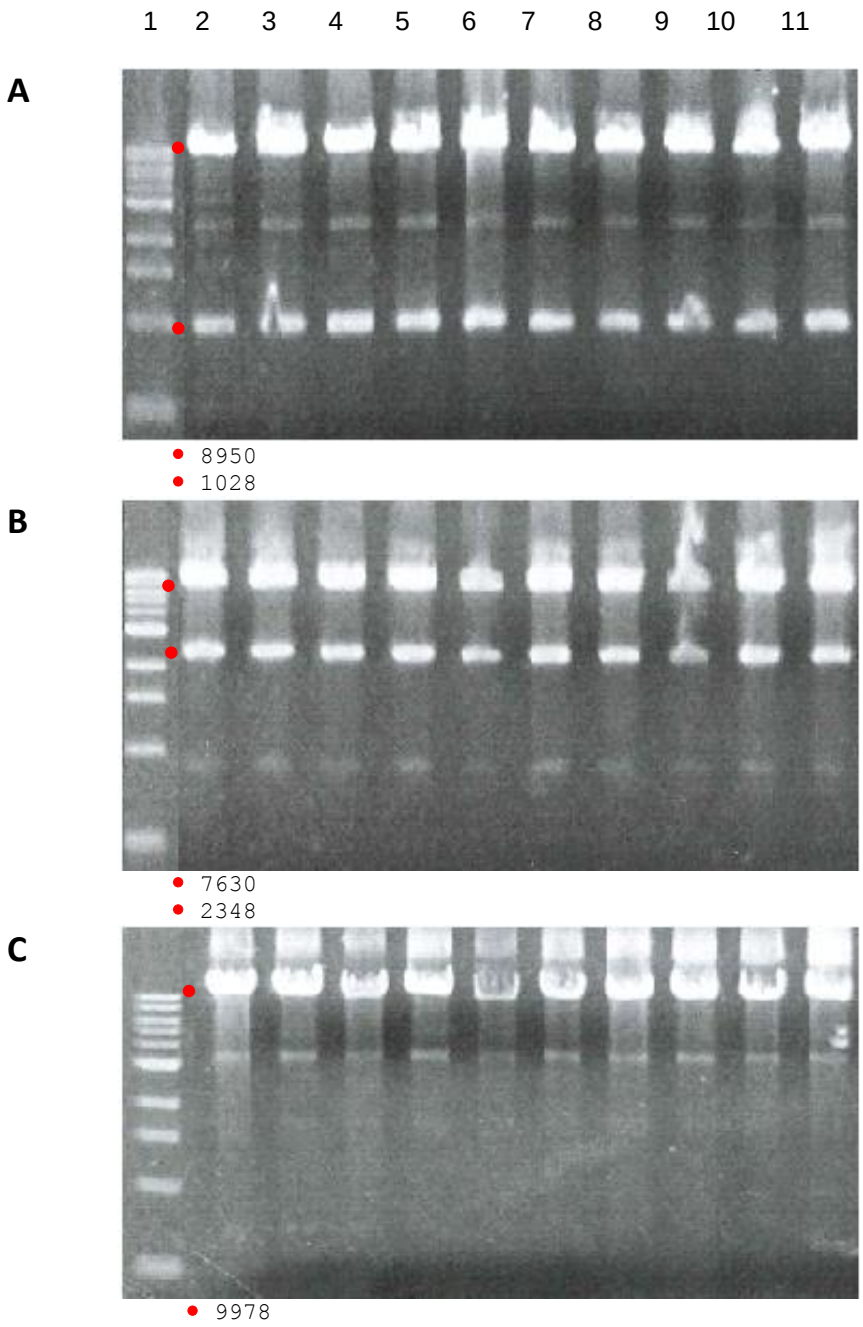


Figure 3.1.13. Agarose gel separation of MA821Qsh_B1_B9 digested with *Hpa*I, *Nar*I and *Nae*I. Analytical agarose gel electrophoresis of MA821Qsh_B1_B9 candidate clones [2–11] digested with **A)** *Hpa*I, **B)** *Nar*I and **C)** *Nae*I. [1] 1-kb DNA ladder. ●: fragments of interest (bp).

3.1.3.5 Construction of the MA821Qsh_B9 vector

The MA821Qsh_B9 vector was generated by excision of the 59bp BCL11A_451-shRNA and the 57bp GFP-shRNA from the 9978bp MA821Q_SF by *Pfl*MI/*Pme*I digestion, following plasmid re-ligation and transformation of *Turbo* competent *E. coli* cells. Candidate clones were digested with *Hpa*I to confirm the presence of the BCL11A_449-shRNA loop, which produced the expected two fragments of 8950bp and 901bp in three cases (**Figure 3.1.14**). These were further digested with *Nar*I and *Nae*I to confirm absence of respective shRNA elements through plasmid linearization (of 9851 bp) as a result of lacking a second recognition site carried on the excised loops (data not shown). One clone was selected for sequencing using the IVS2_shSEQ primer set (**section 2.3**), which allowed read through the flanking restriction sites up to 6bp into the BCL11A_449-shRNA-loop, confirming the absence of the BCL11A_451-shRNA and the GFP-shRNA loops (data not shown).

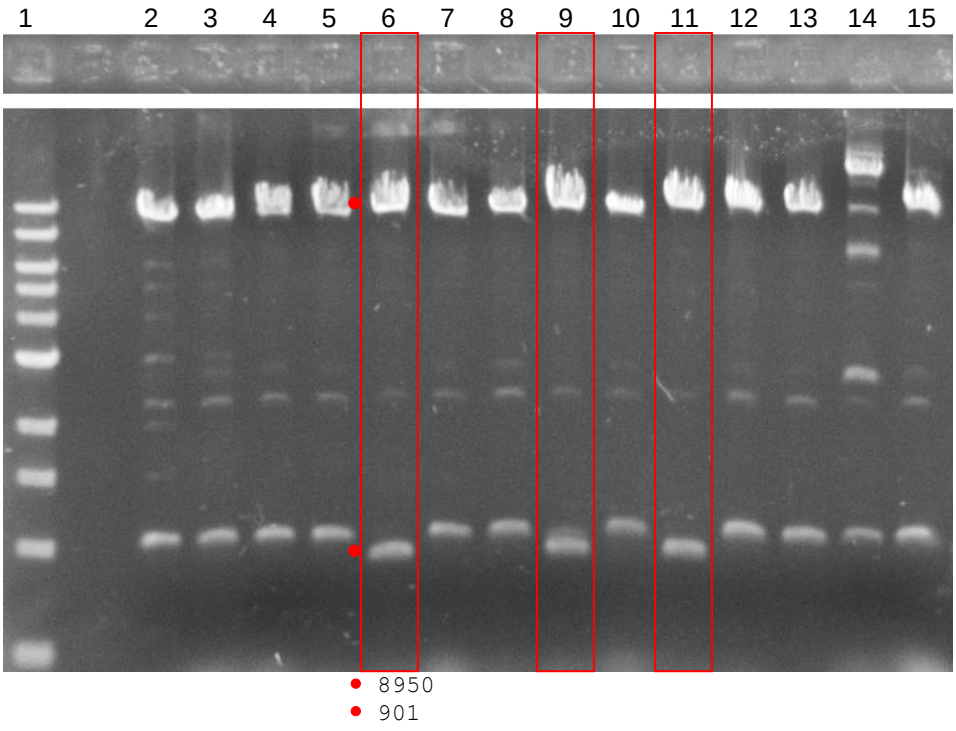


Figure 3.1.14. Agarose gel separation of *Hpa*I-digested MA821Qsh_B9 clones. Analytical agarose gel electrophoresis of MA821Qsh_B9 candidate clones [2–15] digested with *Hpa*I. Correct clones ([6], [9] and [11]) are boxed in red. [1] 1-kb DNA ladder. ●: fragments of interest (bp).

3.1.3.6 Construction of MA821Q_LF

3.1.3.6.1 Cloning LF into the pBS_HBBQ vector

The LF (740bp) synthesised DNA fragment encoding six shRNAs (**Figure 3.1.4**) was provided in the pMA vector flanked by *Hind*III and *Asc*I sites. This prototype pMA_LF plasmid was transformed into XL1 Blue *E coli* bacteria, and candidate clones were confirmed by *Hind*III/*Asc*I digestion, which gave the expected 2397bp and 733bp bands for vector and insert respectively (data not shown). In line with procedures used with pBS_HBBQ_SF, pBS_HBBQ_LF was then produced by re-cloning the pMA_LF 722bp *Pml*I/*Eco*RI fragment into the 5164bp pBS_HBBQ plasmid vector backbone. Putative clones were confirmed by *Hpa*I digestion, which should result in 4670bp and 1212bp fragments for correct clones (**Figure 3.1.15**), while an insert in the reverse orientation would give 5101bp and 781bp fragments. Candidate correct clones were further confirmed by an additional digestion with *Sca*I, resulting in 3624bp and 2258bp fragments. This revealed only one pBS_HBBQ_LF positive clone (**Figure 3.1.16,** lane [6]).

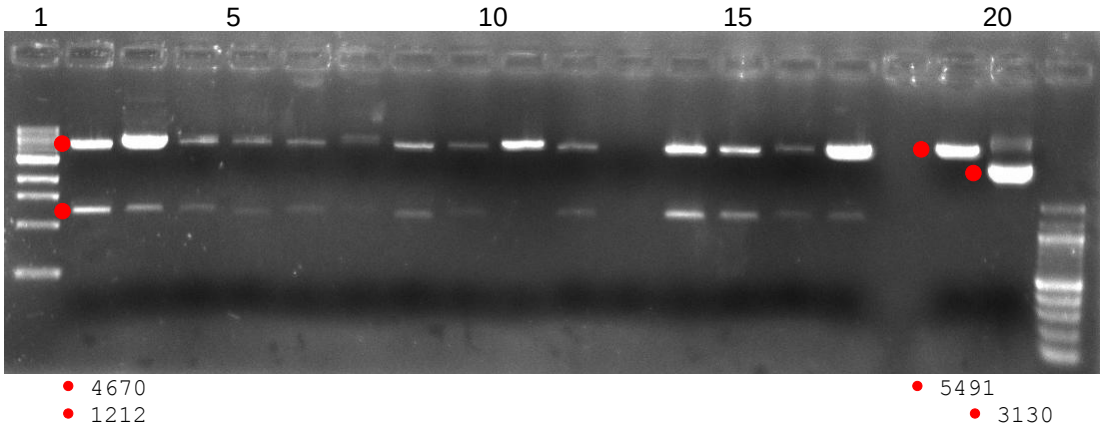


Figure 3.1.15. Agarose gel separation of *Hpa*I-digested pBS_HBBQ_LF clones. Analytical agarose gel electrophoresis of candidate clones of pBS_HBBQ_LF [2–16], pBS_HBBQ [18], and pMA_LF [19] digested with *Hpa*I. [1] 1-kb DNA ladder. [20] 100-bp DNA ladder. ●: fragments of interest (bp).

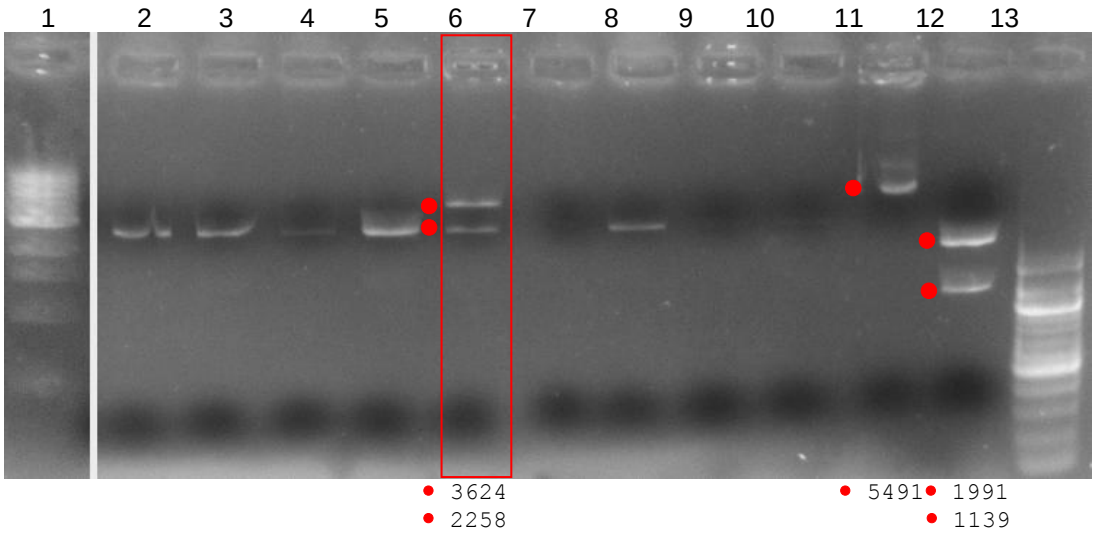


Figure 3.1.16. Agarose gel separation of *Sca*I-digested pBS_HBBQ_LF clones. Analytical agarose gel electrophoresis of candidate clones of pBS_HBBQ_LF [2–10], pBS_HBBQ [11], and pMA_LF [12] digested with *Sca*I. Correct clone ([6]) is boxed in red [1] 1-kb DNA ladder. [13] 100-bp DNA ladder. ●: fragments of interest (bp).

3.1.3.6.2 Cloning LF into the MA821Q vector

Production of the MA821Q_LF plasmid could draw on the availability of MA821Q_SF which, compared to MA821Q, provided more proximal cohesive-end restriction sites for re-cloning. The MA821Q_LF plasmid was thus generated by replacing the 1211bp *Cla*I/*Bsi*WI fragment from the 9978bp MA821Q_SF construct with the 1393bp *Cla*I/*Bsi*WI fragment from the 5882bp pBS_HBBQ_LF. Putative clones following transformation of *E. coli* were analysed by digestion with *Bsr*GI, which despite inefficient digestion revealed two positive MA821Q_LF clones (**Figure 3.1.17,** lanes [13] and [15]), releasing a 385bp fragment, which is equivalent to the size of shRNA-LF from the pBS_HBBQ_LF control, suggesting that these clones contain an insert of the correct size. In order to confirm this preliminary result, both these plasmid clones were further digested with *Hpa*I. This was expected to yield 8950bp and 1212bp fragments for the correct construct. This verified that the first of the two clones was the correct construct (**Figure 3.1.18,** lane [2]), with the second as probably identical with the pBS-HBBQ_LF source plasmid.

In addition to the findings from the restriction digestion analyses, clone identity was confirmed by sequencing the construct across the cloning sites (**Figure 3.1.19**). As before, sequencing did

not pass the outer hairpins of the shRNA cassette, but confirmed its presence, which had been previously sequenced internally including by the manufacturer (GeneArt).

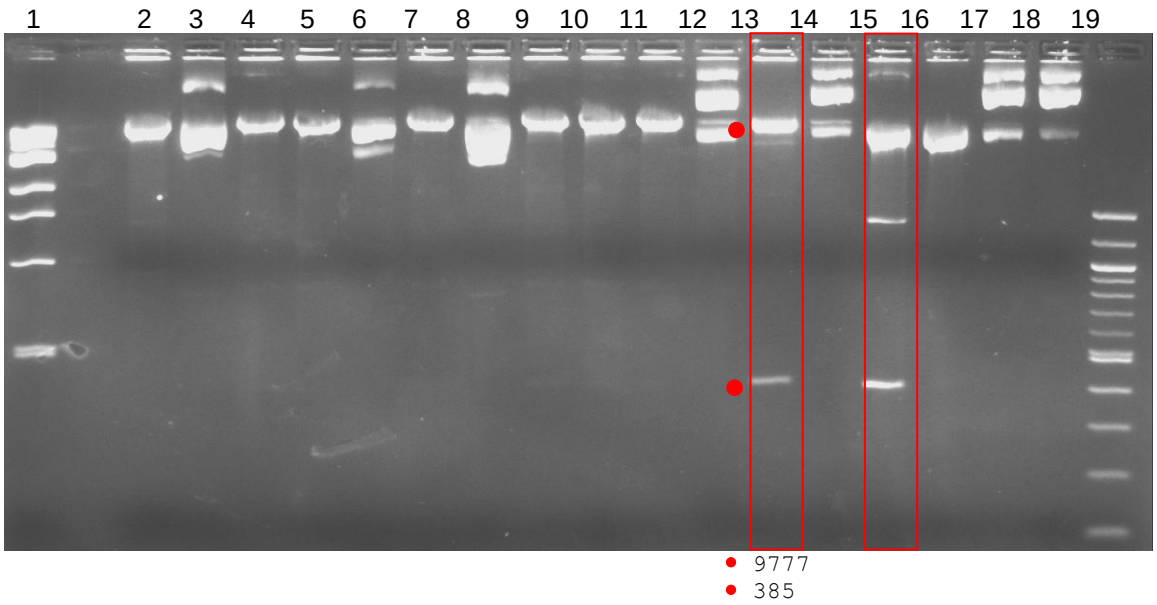


Figure 3.1.17. Agarose gel separation of *Bsr*GI-digested MA821Q_LF clones. Analytical agarose gel electrophoresis of candidate MA821Q_LF clones [2–9] digested with *Bsr*GI. Correct clones ([13], [15]) are boxed in red [1] 1-kb DNA ladder. [9] 100-bp DNA ladder. •: fragments of interest (bp)

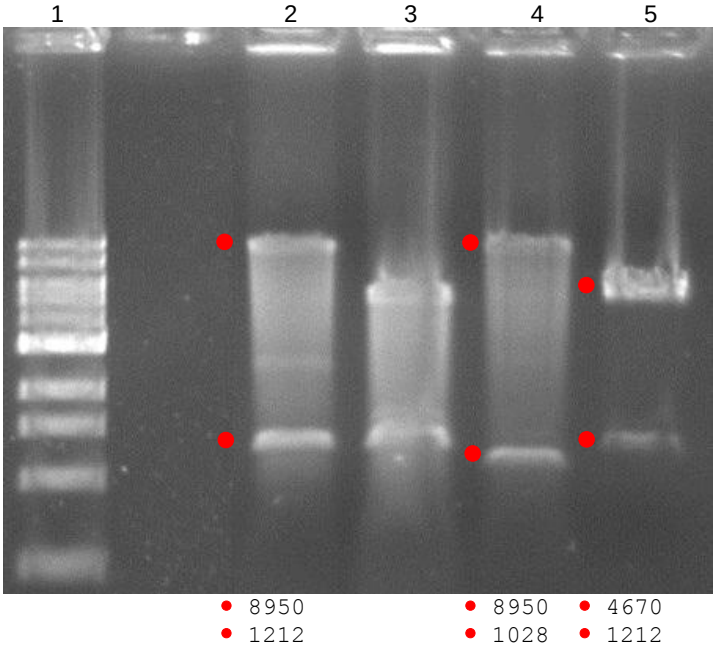


Figure 3.1.18. Agarose gel separation of confirmatory *Hpa*I-digested MA821Q_LF clones. Analytical agarose gel electrophoresis of MA821Q_LF candidate clones [2–3], MA821Q_SF [4], and pBS_HBBQ_LF [5] digested with *Hpa*I. [1] 1kb DNA ladder. •: fragments of interest (bp).

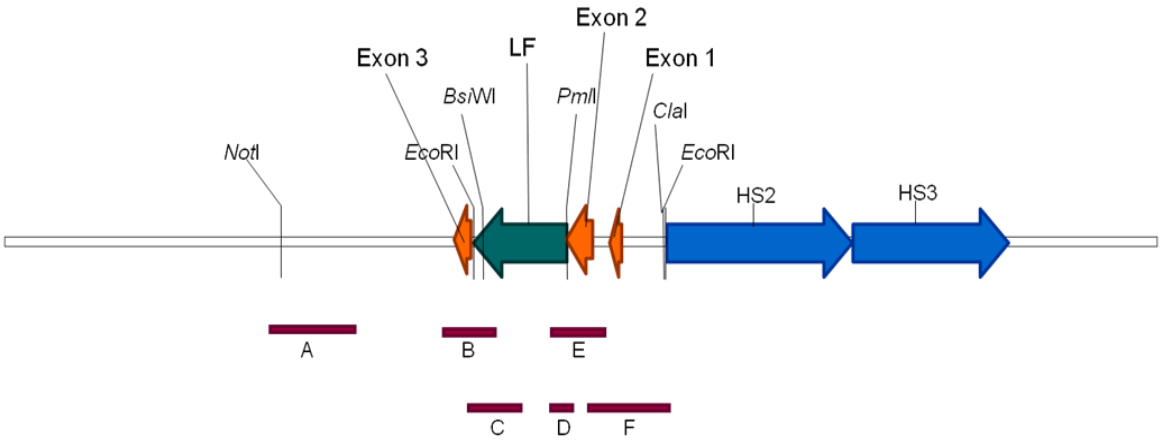


Figure 3.1.19. Confirmatory sequence analysis of the MA821Q_LF plasmid clone. Schematic of sequencing primers for *HBB* in the MA821Q_LF vector. The position of the primers around the *Pml*I/*Eco*RI and *Cla*I/*Bsi*WI cloning sites, and the read length relative to the construct sequence is shown. A: MA821_NotI_FW; B: ARMS 10; C: IVS2_shSEQ_RV, D: IVS2_shSEQ_FW; E: αPCO4_FW, F: MA821_ClaI_RV.

3.1.4 Construction of shRNA-miR expression cassettes

The construction of the shRNA-miR-based vectors followed on from preliminary results suggesting a possibly non-functional shRNA structure and a literature search suggesting the likelihood of having a sub-optimal shRNA design for RNA-pol-II-mediated expression in the modified GLOBE-derived vectors. This prompted the use of artificial shRNA-miR constructs as an alternative strategy to achieve improved siRNA-mediated knockdown over their shRNA-based counterparts [303-305]. An advantage of this approach is the fact that the shRNA is flanked by genomic miRNA sequences that are naturally present in miRNA genes and which support optimal substrate processing. The cloning strategy of the shRNA-miR fragments in the GLOBE vector is illustrated in **Figure 3.1.20**.

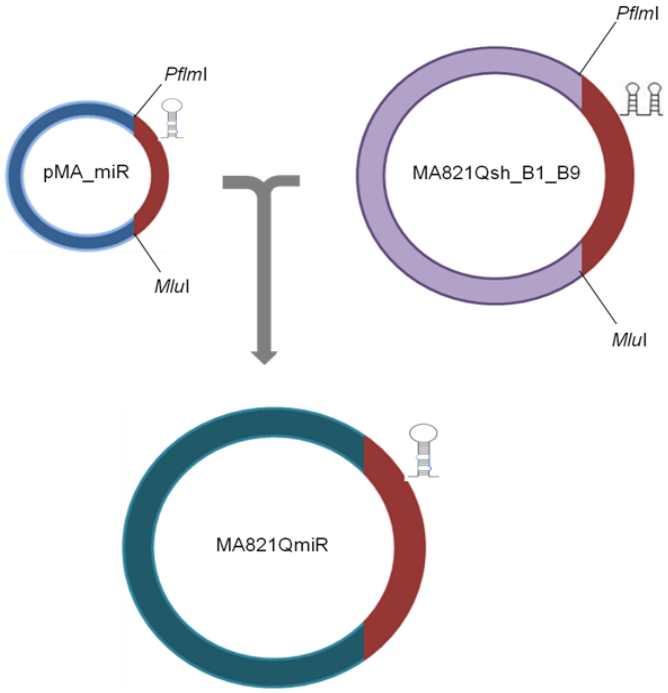


Figure 3.1.20. Cloning strategy for shRNA-miR-expressing LV vector production. The diagram illustrates the cloning of miRNA-encoding fragments into the *HBB* expressing MA821Q vector. The miRNA cassettes were inserted into the MA821Qsh_B1_B9 plasmid, replacing the two standard BCL11A shRNAs, by taking advantage of the additional restriction enzyme sites introduced in the vector backbone during initial modification steps compared to the full-length MA821Q vector.

3.1.4.1 Design and structure of shRNA-miR cassettes

The shRNA closely resembles the authentic pre-miRNA structure, an intermediate in the miRNA biogenesis pathway generated by the cleavage of long primary transcripts (pri-miRNAs) by the RNase III enzyme Drosha [306]. This means that the shRNA is not processed by Drosha. Instead, the shRNA would enter the processing machinery at the level of Dicer. In order to achieve an effective RNAi knockdown, efficient processing by the Dicer/TRBP complex is imperative, which otherwise would increase the likelihood of an aberrant guide/passenger duplex ultimately leading in weak RNAi and induction of off-target effects [209]. To date, the structural requirements for effective Dicer cleavage are not, as yet, fully defined, while its processing sites on the hairpins were reported to vary with respect to the flanking sequences at the end of the stem [227]. To this end, at the moment, shRNA technology uses RNA-pol-III

transcription and follows well defined structural design guidelines for successful shRNA expression.

As an alternative, the use of artificial shRNA-miRs was reported by many to substantially improve knockdown [303, 305, 307-309] and holds great promise over the conventional shRNA design since the miRNA structure serves as a natural substrate in miRNA biogenesis. In our vectors, the shRNA construct is embedded within the intronic region of the *HBB* transgene under a RNA-pol-II-based expression system. The spliced intron containing the artificial shRNA-miR possesses an analogous structure to the endogenous pri-miRNAs (intronic sequences) such that the artificial design would possess an optimal structural configuration for effective processing by both Dorsha and Dicer enzymes [306]. Currently, there are no detailed studies defining the requirements for efficient RNA-pol-II-shRNA expression, especially in terms of hairpin position, stem length and transcription start/termination sites.

Moreover, artificial miRNA-based systems reportedly circumvent the structural deficiencies imposed by RNA-pol-II-dependent shRNA expression and could offer effective knockdown. RNAi technology has rapidly grown into an important method for analyzing gene expression and gene function in many different models, holding great promise for therapeutic gene silencing. Understanding of the endogenous miRNA biogenesis pathway led the way for the development of pooled siRNA libraries for functional genomic screens [310]. Our vector design is based on the second generation of the Hannon-Elledge shRNA libraries, and in particular their improved LV model expressing shRNAs from the RNA-pol-II promoter CMV, the so-called pGIPZ vector [311]. Given the immense therapeutic potential of the RNAi technology platform, this library became publically available and was further developed to offer targeting of the majority of human genes (GIPZ & TRIPZ shRNAmir lentivector expression systems; *Open Biosystems; GE Dharmacon, Lafayette, CO*). The prototype model built a shRNA-miR cassette within the pSM2 vector under the control of the RNA-pol-III U6 promoter and employed the miRNA miR-30 scaffold to ensure effective processing and knockdown. The shRNA-miR constructs retained the flanking sequences of pre-miR-30 and replaced the native miRNA stem with synthetic siRNA sequences separated by the 15nt miR-30 loop [304, 305, 312]. The miR-30 miRNA was isolated from the human cell line HeLa during initial studies to identify novel short RNAs involved in post-

transcriptional regulation. This was also the first time to demonstrate successful expression of the miR-30 precursor from the RNA-pol-II CMV promoter [313]. Based on these results, subsequent studies worked on miR-30 to decipher its structural composition and map its RNAi processing sites (**Figure 3.1.21**) [314-318]. Additionally, a recent study proposed that precise Dicer cleavage follows a 'loop-counting rule' and identified compliance with the miR-30 natural loop configuration [209]. Thus taking advantage of existing studies defining the structural requirements for miR-30 processing and function in human cells, we designed constructs expressing gene-specific shRNA target sequences from within the miR-30 backbone (**Table 3.1.2**).

Figure 3.1.21. Design of the human miRNA miR-30. The Drosha and Dicer processing sites are shown. Adapted from [306].

Hairpin structure	Sequence (5’ to 3’)	Ref.
5' miR30 (vector_portion)	TGTTTGAATGAGGCTTCAGTACTTTACAGAATCGTTGCCTGCACATCTTGGAAACACTTG CTGGGATTACTTCTTCAGGTTAACCCAACAGAAGG	*pGIPZ* vector
5’ precursor miR-30 arm	CTAAAGAAGGTATATTGCTGTTGACAGTGAGCG	[319]
shRNA stem (sense)	see **Table 3.1.1**	
miR30 Loop	CTGTGAAGCCACAGATGGG	[306]
shRNA stem (anti-sense)	see **Table 3.1.1**	
3’ precursor miR-30 arm	TGCCTACTGCCTCGGACTTCAAGGGG	[319]
3’ miR-30 (vector portion)	AGGGGCTACTTTAGGAGCAATTATCTTGTTTACTAAAACTGAATACCTTGCTATCTCTTTG ATACATTTTTACAAAGCTGAATTAAAATGGTATAAATTAAATCACT	*pGIPZ* vector

Table 3.1.2. Design of miRNA-shRNA hairpin based on the human miR30α scaffold. The miR-30 backbone of the shRNA-miR hairpin includes precursor miR-30α arm sequences with an extended 95–107nt flanking sequence to resemble the natural primary miRNA structure necessary for correct processing. The shRNA sense and anti-sense strands are connected by the natural miR-30α loop sequence. Blast sequence alignment of the modified miR-30 scaffold to the endogenous human pri-miRNA miR-30 backbone lacking its duplex stem (NC_018917) identified a 3nt mismatch at the 5’ pre-miRNA arm. A 100% sequence similarity was reached after inclusion of the CTA trinucleotide at the 5’ end of the recommended 5’ pre-miRNA sequence. The shRNA-miR structure was designed to carry an *Mlu*I site at the 5’end and a *Pfl*MI site at the 3’end to facilitate cloning in the MA821Qsh_B1_B9 vector, as well as *a Bsr*GI site at both ends which allows cloning in any of the MA821Q vectors modified to carry the shRNA expressing cassette.

3.1.4.2 Construction of MA821Qmir_B9 and MA821Qmir_B1

The miRNA expression cassettes (346bp) were provided in the pMA vector, flanked by *Sfi*I sites. To construct the MA821Qmir_B9 and MA821Qmir_B1 vectors, the miRNA target sequences were excised by *Pfl*MI/*Mlu*I digestion as a 329bp fragment and ligated into the corresponding sites of the 9921bp MA821Qsh_B1_B9 vector. Following transformation into *Turbo* competent cells, correct clones were confirmed by digestion with *Sca*I. Resolution of products by agarose gel electrophoresis identified one out of the screened clones for each MA821Q-miRNA construct as being correct (**Figure 3.1.22,** lanes [8] and [10]). Given the similar size of digestion products (of 5367 bp and 4758 bp) after *Sca*I digestion (**Figure 3.1.22**), the insertion of insert in the correct orientation was confirmed by *Hpa*I digestion, which should give 9007bp and 1118bp fragments, whilst an insert in the incorrect reverse orientation would result in 9287bp and 838bp fragments. The results confirm that both of the candidate clones identified following initial *Sca*I digestion (**Figure 3.1.22**) contained the miRNA expression cassette in the correct orientation (**Figure 3.1.23,** lanes [3] and [4]). These two positive clones were verified by sequencing the constructs using the T87Qx2x3_FW primer which read through the cloning sites across the 329bp miRNA hairpin (data not shown).

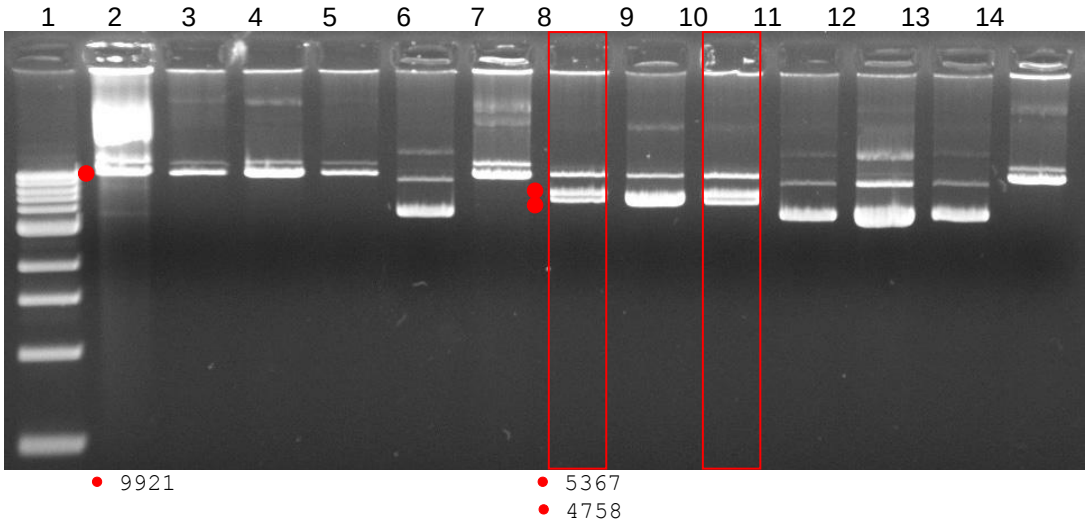


Figure 3.1.22. Agarose gel separation of *Sca*I-digested MA821Qmir_B1 and MA821Qmir_B9 clones. Analytical agarose gel electrophoresis of MA821Qsh_B1_B9 candidate clones [2], MA821Qmir-B1 [3–8], and MA821Qmir-B9 [9–14], digested with *Sca*I. Correct clones ([8], [10]) are boxed in red. [1] 1-kb DNA ladder. •: fragment of interest (bp).

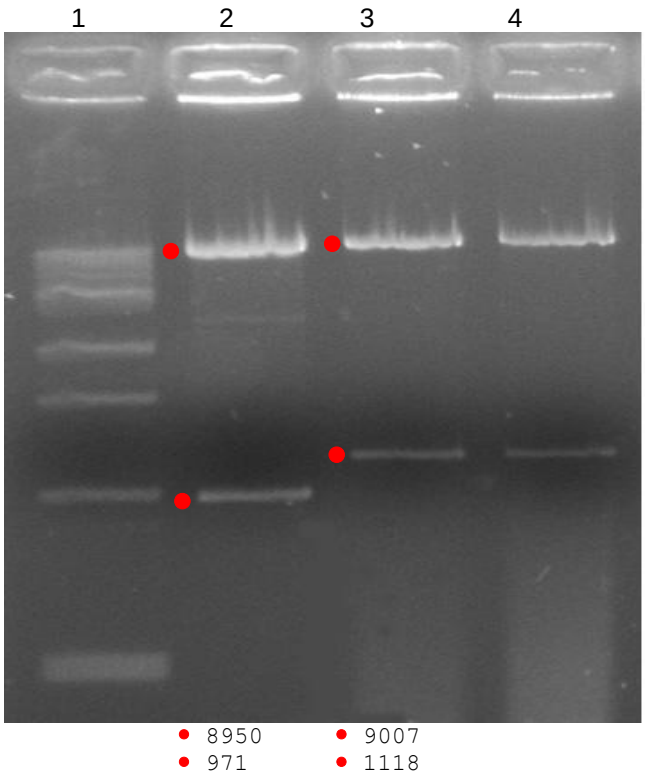


Figure 3.1.23. Agarose gel separation of *Hpa*I-digested MA821Qmir_B1 and MA821Qmir_B9 clones. Analytical agrose gel electrophoresis of MA821Qsh_B1_B9 [2], MA821Qmir-B9 [3], and MA821Qmir-B1 [4], digested with *Hpa*I. [1] 1-kb DNA ladder. •: fragment of interest (bp).

3.1.4.3 Construction of MA821Qmir_Scr

In line with procedures for MA821Qmir_BCL11A vectors, the control MA821Qmir_Scramble vector was then produced by re-cloning the pMAmir_Scr 329bp *Pfl*MI/*Mlu*I fragment into the 9803bp MA821Qsh_B1_B9 plasmid vector backbone. Plasmids from putative clones following transformation into *E. coli* were confirmed by *Sca*I digestion, that should result in a 5367bp and 4758bp fragments. A total of 6 clones gave the correct banding pattern (**Figure 3.1.24**). Of the candidate clones, two (**Figure 3.1.24**, lanes [3] and [9]) were selected for further analysis by digestion with *Hpa*I. This resulted in the expected 9007bp and 1118bp fragments confirming that indeed both clones were correct (**Figure 3.1.25,** lanes [3] and [4]). These positive clones were verified by sequencing using the T87Qx2x3_FW and the IVS_SEQ_RV primers (**section 2.3**), which read across the cloning sites into the hairpin covering 127bp at the 5' end and 142bp at the 3' end of the inserted cassette (data not shown).

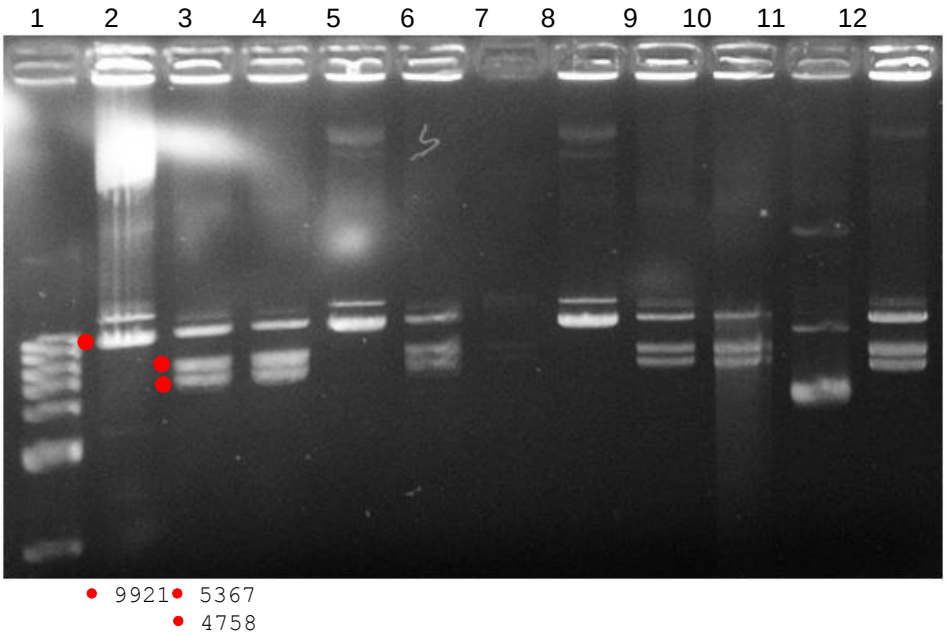


Figure 3.1.24. Agarose gel separation of *Sca*I-digested MA821Qmir_Scr clones. Analytical agarose gel electrophoresis of MA821Qsh_B1_B9 [2], and MA821Qmir-Sc [3–12] digested with Sca*I*. Correct clones are ([3, 4, 6, 9, 10,12]). [1] 1-kb DNA ladder. •: fragment of interest (bp).

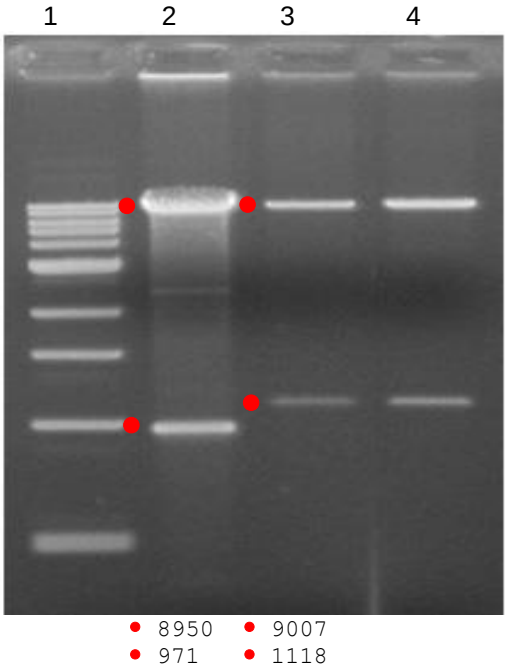


Figure 3.1.25. Agarose gel separation of *Hpa*I-digested MA821Qmir_Scr clones. Analytical agarose gel electrophoresis of MA821Qsh_B1_B9 [2], MA821Qmir-Scr [3–4], digested with *Hpa*I. [1] 1-kb DNA ladder. •: fragment of interest (bp).

3.1.5 Construction of pLKO vectors

The siRNA potency of the shRNA hairpins was investigated by employing a standard U6 promoter-based expression platform. Control vectors were constructed expressing the shRNAs under the U6 RNA-pol-III promoter [320], which is known to work in mammalian cells and, in contrast to other RNA-pol-III promoters, does not require internal elements downstream of +1 site for efficient transcription [321]. RNA-pol-III-directed shRNA expression has proved to be an extremely valuable system for targeting gene knockown and studying gene function, because the transcription initiation and termination sites are well defined. In particular, a key feature of the shRNA structure is the inclusion of a stretch of five thymine (T) residues in the terminator sequence forming a 2nt T 3' overhang in the transcribed product, which is essential for Dicer processing [322, 323], especially, given the fact that nuclear Drosha cleavage is not taking place. Based on the constitutive activity of RNA-pol-III promoters, together with their efficient structural features, the TRC lentiviral shRNA library was constructed to facilitate high-throughput screening across multiple cell types. The vectors are based on the lentiviral pLKO.1 backbone and include sequence-validated shRNAs expressed from the human U6 promoter. Currently, the library contains over 100,000 vectors harbouring shRNA constructs against 12,000 human and 10,000 mouse genes [324]. Our vector design follows the TRC guidelines and modifies the publically available pLKO.1 stuffer plasmid (*Addgene*) for U6-directed shRNA expression (**Figure 3.1.26**). The siRNA target sequence of the shRNA is identical to the one cloned and expressed under the RNA-pol-II human *HBB* promoter of the MA821Q vector.

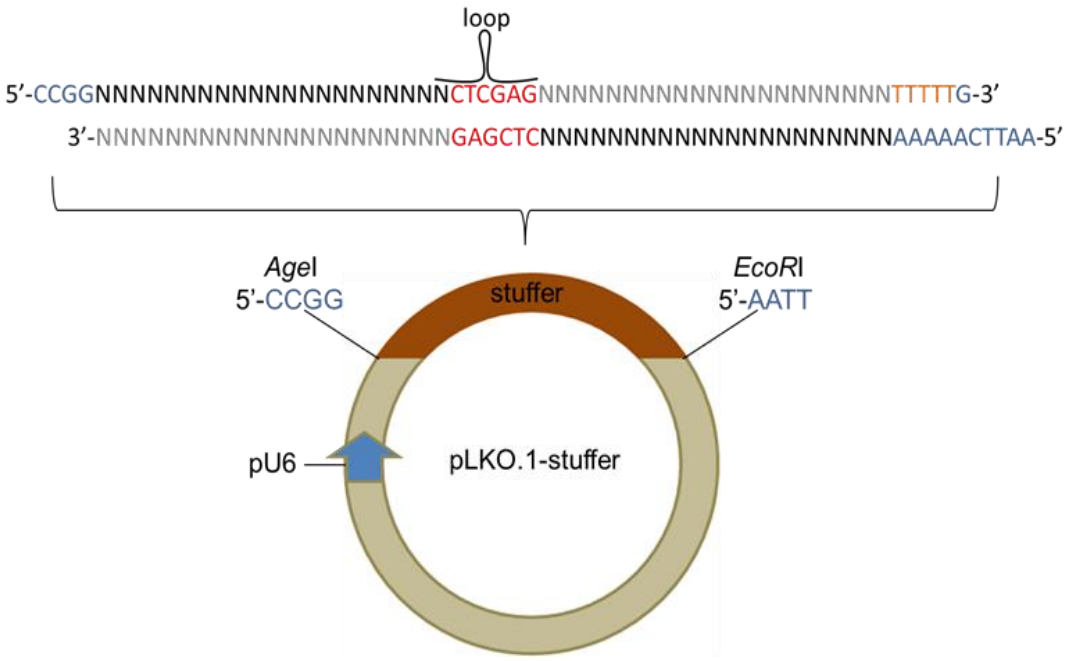


Figure 3.1.26. Cloning strategy for shRNA expression under the control of the human RNA-pol-III U6 snRNA promoter from within the lentiviral pLKO.1 TRC vector. The shRNA design contains a sense strand 21nt sequence of the target transcript (N in black) separated by a short palindromic spacer (TRC-validated *Xho*I site in red) from the 21nt antisense strand complementary to the sense bases (N in grey), and ends with a five thymidine base stretch (T_5 in orange) as an RNA-pol-III transcriptional stop signal. The end bases (in blue) of the shRNA oligonucleotides are essential for cloning into the pLKO.1 vector. These bases form compatible cloning sites to the *Eco*RI and *Age*I ends flanking the 1.9kb stuffer sequence in the pLKO.1 vector, which is replaced to produce a plasmid expressing the shRNA of interest.

3.1.5.1 Construction of pLKOsh_GFP

The pLKOsh_G vectors were produced to quickly determine the silencing efficiency of the shRNA design using flow cytometry and to correlate the intensity of GFP fluorescence to shRNA-mediated GFP knockdown. A total of three GFP-specific shRNAs were designed, all with an identical siRNA target sequence [302] but a different loop structure. The GFP-shRNA_a hairpin was designed as an exact replica of the shRNA expressed under the *HBB* promoter of the MA821Qsh_G vector, and would inform on the functionality of the βIVS-II insertion site. The GFP-shRNA_b was designed to include the 9bp spacer of the published GFP-specific shRNA from which the siRNA GFP-target sequence was reproduced, and would serve to assess the

functionality of our vector system considering that effective knockdown was previously demonstrated with a RNA-pol-III H1 promoter-based viral system [302]. The GFP-shRNA_c hairpin included the 6bp loop sequence recommended by the shRNA design guidelines in the TRC pLKO manuscript [324] and which was also applied for all other gene-specific shRNA designs, and would assess the effect of the loop sequence/structure on shRNA processing.

Two oligonucleotides complimentary to each other were designed and synthesized for the construction of the pLKOsh_G vectors. The oligonucleotide DNA template sequence is shown below:

5' oligo (FW):
5'-CCGGGCAAGCTGACCCTGAAGTTC  GAACTTCAGGGTCAGCTTGCTTTTTG-3'

3' oligo (RV):
5'-AATTCAAAAAGCAAGCTGACCCTGAAGTTC  GAACTTCAGGGTCAGCTTGC-3'

The siRNA target sequence for the GFP mRNA is 20nt in length, shown in black, and common for all GFP-specific shRNA hairpins. The RNA-pol-III terminator is shown in orange and the sticky ends attached to the 5' end complementary with *Age*I and *Eco*RI digested ends are shown in blue. The loop structure and its location are schematically annotated and include the following sequences for each designated hairpin:

plasmid	hairpin	loop (5' – 3')
pLKOsh_Ga	GFP-shRNA_a	TTCGCCGGC
pLKOsh_Gb	GFP-shRNA_b	TTCAAGAGA
pLKOsh_Gc	GFP-shRNA_c	CTCGAG

To construct the pLKOsh-G plasmids, the stuffer sequence from the 8901bp pLKO.1_stuffer plasmid was excised as an 1875bp *Eco*RI/*Age*I fragment and the gel-purified 7026bp vector DNA was ligated with the short hairpin sequence followed by transformation of *Turbo* competent cells. Plasmids isolated from resulting bacterial clones were analysed by digestion with *Ngo*MIV,

*Xmn*I and *Xho*I, which are restriction sites encoded in the loop sequence of GFP-shRNA_a, GFP-shRNA_b and GFP-shRNA_c hairpins, respectively. Correct clones were identified by digestion with *Ngo*MIV, which produced two products (4655bp and 2430bp), with *Xmn*I producing three products (3084bp, 2518bp and 1483bp), and with *Xho*I producing four products (6556bp, 294bp, 190bp and 42bp, with the last not visible following agarose gel resolution) (**Figure 3.1.27**). Insertion of the shRNA sequence was also verified by sequencing using the pLKO.1 SEQ primer, which read across the full length of the insert, which included the cloning sites and the hairpin structure, for all three pLKOsh_G vectors (data not shown).

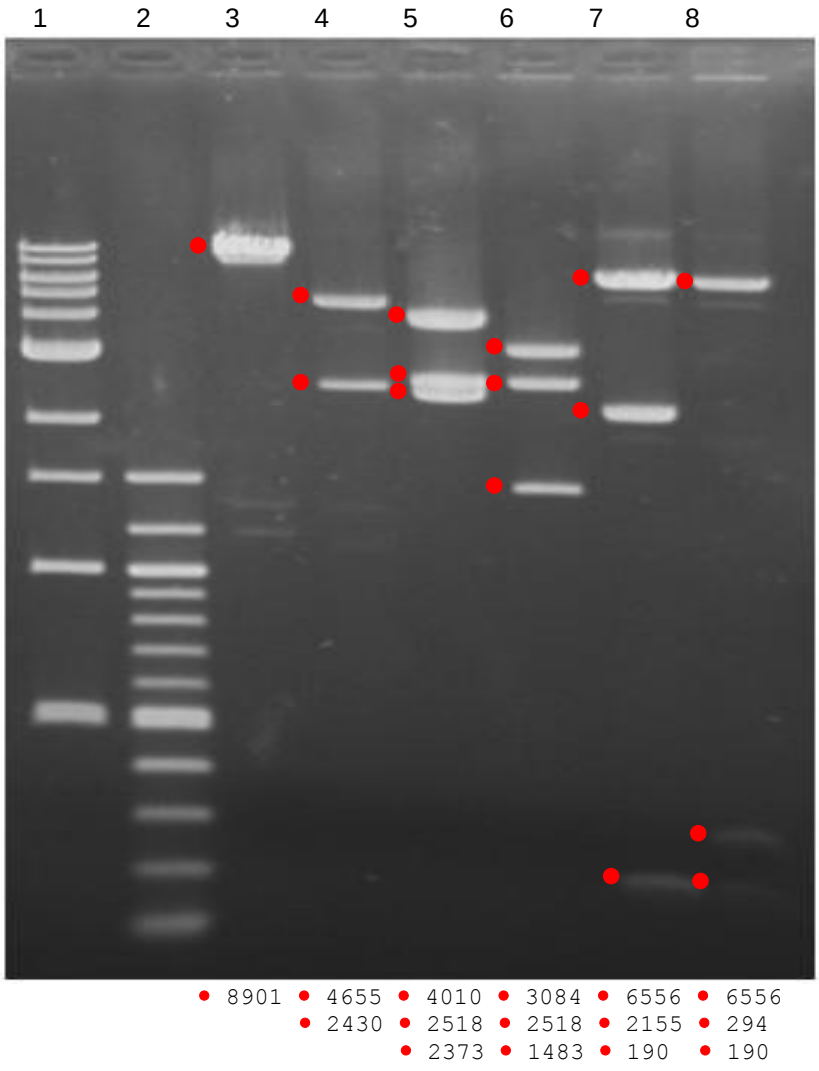


Figure 3.1.27. Agarose gel separation of *Ngo*MIV, *Xmn*I, and *Xho*I-digested pLKO.1 stuffer and pLKOsh-G clones. Analytical agarose gel electrophoresis of pLKO.1 stuffer digested with *Ngo*MIV [3], *Xmn*I [5], *Xho*I [7], pLKOsh-Ga digested with *Ngo*MIV [4], pLKOsh-Gb digested with *Xmn*I [6], and pLKOsh-Gc digested with *Xho*I [8]. [1] 1-kb DNA ladder, [2] 100-bp ladder. •: fragments of interest (bp).

3.1.5.2 Construction of pLKOsh_BCL11A

In line with procedures for pLKOsh_GFP, the pLKOsh_BCL11A plasmids were produced, providing a useful tool to demonstrate functional knockdown of the target genes, thus serving as positive controls in experiments investigating the temporal expression of BCL11A-specific shRNA under the control of the late-erythroid *HBB* promoter from the MA821Qsh_BCL11A vector. Our investigation uses two siRNA target sequences for the human BCL11A mRNA, such that, for each siRNA two sets of pLKOsh_BCL11A clones were produced, one with the 6nt loop sequence (*Xho*I) of the publically validated shRNA and the other with that used in the MA821Q-derived plasmids. Primer oligonucleotide sequences are shown below:

BCL11A(shRNA449) hairpin:

5’ oligo (FW):
5’-CCGGCGCACAGAACACTCATGGATT AATCCATGAGTGTTCTGTGCGTTTTTG-3’

3’ oligo (RV):
5’-AATTCAAAAACGCACAGAACACTCATGGATT AATCCATGAGTGTTCTGTGCG-3’

BCL11A(shRNA451) hairpin:

5’ oligo (FW):
5’-CCGGCCAGAGGATGACGATTGTTTA TAAACAATCGTCATCCTCTGGTTTTTG-3’

3’ oligo (RV):
5’-AATTCAAAAACCAGAGGATGACGATTGTTTA TAAACAATCGTCATCCTCTGG-3’

And loop sequences for each clone are as follows:

hairpin	clone	Loop (5’ – 3’)
BCL11A_449-shRNA	pLKOsh_B9	GTTAAC
	7-1a	CTCGAG
BCL11A_451-shRNA	pLKOsh_B1	GGCGCC
	8-1	CTCGAG

In order to construct the pLKOsh-BCL11A plasmids, the 1875bp stuffer sequence in the 8901bp pLKO.1 vector was excised by *Eco*RI/*Age*I digestion and was replaced with the short shRNA oligonucleotides via compatible restriction end ligation, followed by transformation of *Turbo* (for pLKOsh_B9 and pLKOsh_B1) and XL1 blue (for 7-1a and 8-1) *E. coli* cells. Clones were analysed by digestion with *Hpa*I, *Nar*I and *Xho*I for which recognition sequences are encoded in the loop of the hairpins. Correct pLKOsh_B9 clones produced a linear product (of 7074bp) after *Hpa*I digestion, pLKOsh_B1 clones produced five products (4281bp, 1690bp, 654bp, 318bp and 131bp) after *Nar*I digestion, 7-1a clones produced four products (6556bp, 295bp, 190bp, and 43bp) and 8-1 clones produced four products (6556bp, 295bp, 190bp, and 43bp) after *Xho*I digestion (**Figure 3.1.28**). Positive clones were verified by sequencing using the pLKO.1 SEQ primer which covered the whole insert, including cloning sites (data not shown).

3.1.5.3 Validation of the pLKOsh_Scramble

The shRNA scrambled pLKO.1 vector was obtained from *Addgene* and acts as an siRNA negative control to investigate sequence-independent off-target effects triggered by the introduction of small RNAs in the cellular system investigated, as well as, possible saturation of the RNAi machinery by the administered siRNA duplexes, which would out-compete endogenous miRNAs and possibly disturb normal gene regulation.

The scrambled siRNA duplex control sequence is shown below:

5’- CCTAAGGTTAAGTCGCCCTCGCTCGAGCGAGGGCGACTTAACCTTAGGTTTTT-3’

The scrambled siRNA sequence is shown in black and includes the T_5 termination signal at the 3’ end which is depicted in orange. The two inverted repeats are separated by a 6bp palindromic spacer (*Xho*I) which is highlighted and schematically illustrated by a loop. Confirmatory digests involved cutting through the loop with *Xho*I, which also cuts the vector backbone at three sites, producing four products (6556bp, 296bp, 190bp, and 43bp), the

smallest of which is not visible following resolution on an agarose gel due to its small size (**Figure 3.1.28,** lane [7]). The plasmid was sequenced by the manufacturer.

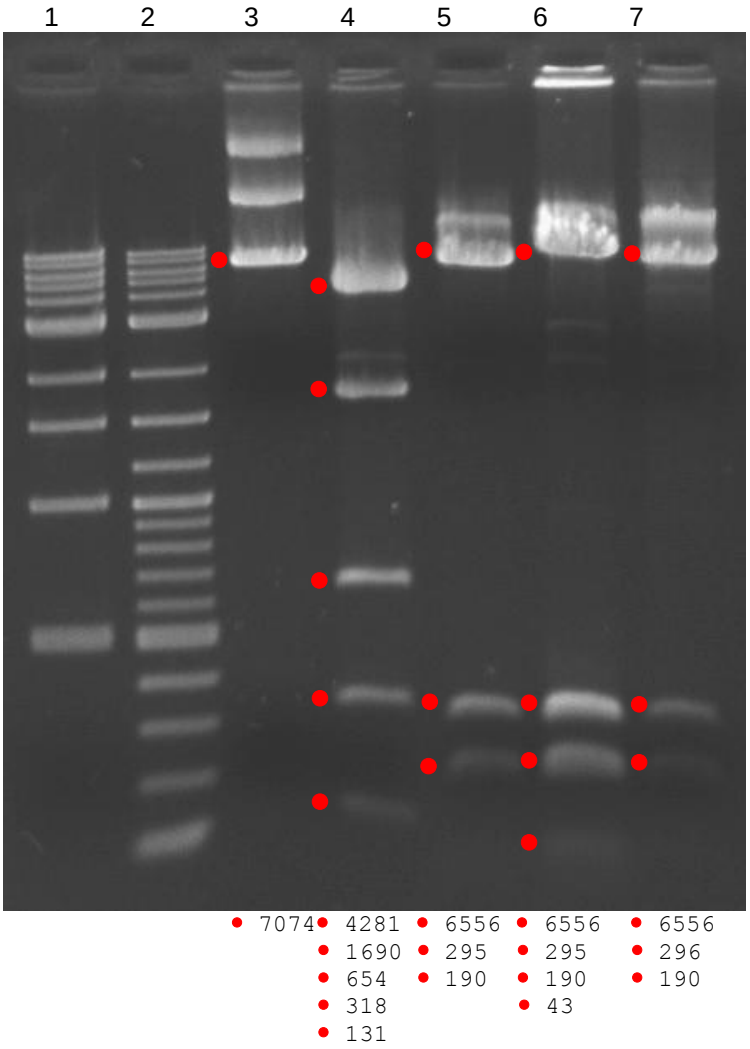


Figure 3.1.28. Agarose gel separation of *Hpa*I, *Nar*I, and *Xho*I-digested pLKO-shRNA clones. Analytical agarose gel electrophoresis of pLKOsh_B9 digested with *Hpa*I [3], pLKOsh_B1 digested with *Nar*I [4], 7-1a [5], 8-1 [6] and pLKOsh_Scr [7] digested with *Xho*I. [1] 1-kb DNA ladder, [2] 100-bp ladder. •: fragments of interest.

3.1.6 Lentiviral vector production

LVs were produced using standard methods by co-transfection of HEK 293T cells with the transfer-vector and packaging/envelope plasmid constructs (see **Figure 2.8**). Challenges were faced resulting from difficulties with low yields of the plasmid DNA required for vector production. Super Broth instead of standard LB medium and larger culture volumes for bacterial growth overcame this difficulty from apparent insert-related growth retardation or toxicity, and increased DNA yields sufficiently. In order to increase viral titres, the production process was scaled-up by using a larger number of plates of HEK 293T cells for transfection, followed by LV concentration (X350) as described (**section 2.13.2**) using protocols, the efficiency of which had previously been tested and confirmed in control experiments with LV-PGK-GFP vectors. Vectors were used to transduce HEL cells (**section 2.13.3**), followed by quantification of the biological titre by qPCR (**section 2.13.4.2**), which achieved >10^8 TU/mL for a series of different transfer vectors (**Figure 3.1.29**).

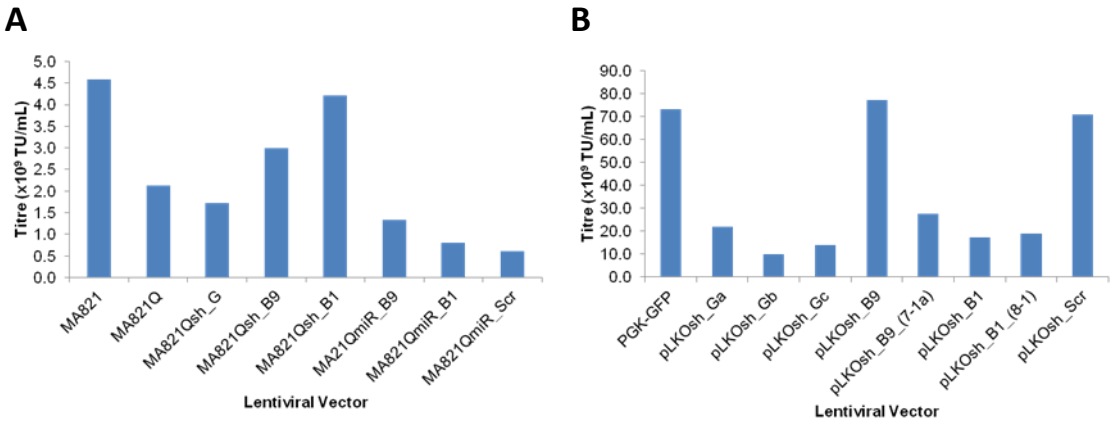


Figure 3.1.29. LV titre quantification. LVs were produced using standard protocols involving co-transfection of HEK 293T cells with transfer-vector and packaging/envelope plasmids followed by concentration of culture supernatant by centrifugation (20,000 g / 4 hours / 4 °C), yielding titres (x10^9 TU/mL) for **A)** the *HBB*-expressing vectors MA821, MA821Q and their shRNA-based permutations, as well as **B)** the PGK-GFP and the shRNA-encoding U6-bsased pLKO.1 vectors. Titres were assessed by qPCR at 3-weeks post-transduction of HEL cells.

3.1.7 Discussion

This chapter describes the production of *HBB* expressing vectors carrying shRNAs which, upon expression, are aimed to target negative regulators of *HBG* expression, namely, BCL11A, SOX6, and KLF1. The basic strategy involved modification of the minimal βLCR-HBB cassette in the GLOBE LV [156, 188]. Preliminary studies demonstrated the functionality and therapeutic safety of GLOBE in relevant *in vitro* and *in vivo* pre-clinical assay systems [156, 188]. However, low vector copies (≤1 VCN) were insufficient to raise HbA to the levels necessary to cure β-thalassaemia in $\beta^{0/0}$ cells. Also, as no other extant vector system had achieved transfusion independence in $\beta^{0/0}$ cells, we devised a strategy to develop novel vectors with improved therapeutic efficacy for β-thalassaemia stem cell based gene therapy.

A synthetic DNA fragment encoding multiple shRNAs, each flanked by a unique dual restriction site, was inserted within βIVS-II of a modified GLOBE vector carrying the T87Q codon change (MA821Q) allowing ease of detection of expression from this LV. The shRNA insertion position was close to that described by and successfully used by the Sadelain group [224] to express an anti-sickling HBB-specific shRNA from βIVS-II of an *HBG* transgene driven by the RNA-pol-II *HBB* promoter (**Figure 1.14**). Difficulties in the synthesis of the multi-shRNA-encoding fragments prompted the design and synthesis of a smaller construct in order to expedite production of an initial range of LVs. The smaller construct, containing two BCL11A shRNAs and one GFP shRNA, was swiftly cloned into the intermediate pBS_HBBQ plasmid, whereas sub-cloning into the MA821Q LV transfer plasmid posed a challenge. Approaches to enhance cloning efficiencies included new restriction enzyme stocks, alternative restriction digests, de-phosporylation of vector backbone, adding PEG6000 to enhance ligation efficiencies, and different strains and preparations of *E. coli* competent cells to enhance transformation efficiencies. A switch to super-competent bacteria and replacing blunt-end with alternative, sticky-end-cutting restriction enzymes overcame the cloning bottleneck into the 9kb LV transfer plasmid. Subsequent cloning steps involved the production of a series of LVs carrying subsets of SF-encoded shRNAs. Cloning of the larger multi-shRNA cassette (after its delayed synthesis) into the MA821Q transfer plasmid by replacing the existing smaller cassette with a larger, 385bp fragment was unproblematic. The challenges faced with plasmid production, which significantly

prolonged assessment of the functionality of the vector system, suspended the synthesis of plasmids with subsets of LF-encoded shRNAs. In addition, BCL11A and GFP shRNAs were cloned into the pLKO.1 vector, under the control of the constitutive RNA-pol-III U6 promoter, to test the knockdown potency of the candidate hairpin design. A total of two sets of pLKO.1 shRNA clones were produced, one with the 6nt loop sequence of publically validated shRNAs and the other with that used in the MA821Q-derived plasmids. Furthermore, an additional set of MA821Q-derived plasmids was synthesized late in the project, expressing the shRNA constructs as human miRNA miR-30 primary transcripts. Individual shRNA-miR constructs were designed for the BCL11A-specific siRNA and control scramble-shRNA sequences used in the simple hairpin structures and cloned into the MA821Q transfer plasmid for RNA-pol-II-mediated expression. Plasmid vectors were confirmed by restriction enzyme digest (**section 2.11.1**) and by sequencing (**section 2.3**) after cleaving hairpin structures at the central loop, which for each hairpin sequence represents a unique or double-cutter restriction site within the transfer-vector plasmid.

In summary, a series of different plasmid vectors were created to facilitate production of therapeutic *HBB*-encoding LVs and modified versions encoding shRNA or shRNA-miR strutures, as well as control pLKO-based LVs for candidate shRNAs. Our production/concentration protocols led to high-titre LVs of $\geq 10^8$ TU/mL, which is required for efficient transduction, as well as time- and cost-savings. However, we noted that vectors with a larger insert size (e.g., vectors carrying the *HBB* promoter/βLCR cassette) resulted in lower titres (TU/mL) presumably due to limitations in the viral genome size that can be packaged into a functional virus [174]. A summary of the vectors generated is provided in **Table 3.1.3**.

Vector name	Vector annotation
MA821 (aka GLOBE)	*HBB* transgene
MA821Q	HBB^{T87Q} transgene
MA821Qsh_G	HBB^{T87Q} transgene-based GFP shRNA
MA821Qsh_B9	HBB^{T87Q} transgene-based BCL11A shRNA(449)
MA821Qsh_B1	HBB^{T87Q} transgene-based BCL11A shRNA(451)
MA821QmiR_B9	HBB^{T87Q} transgene-based BCL11A shRNA-miR(449)
MA821QmiR_B1	HBB^{T87Q} transgene-based BCL11A shRNA-miR(451)
MA821QmiR_Scr	HBB^{T87Q} transgene-based Scramble shRNA-miR
pLKOsh_G	RNA-pol-III U6-driven GFP shRNA
pLKOsh_B9	RNA-pol-III U6-driven BCL11A shRNA(449)
pLKOsh_B1	RNA-pol-III U6-driven BCL11A shRNA(451)
pLKOsh_Scr	RNA-pol-III U6-driven Scramble shRNA

Table 3.1.3. Annotation of the lentiviral plasmid vectors generated and used in this study.

3.2 Functional evaluation of GFP-specific shRNAs in mammalian cells

3.2.0 Aims

- To produce immortal GFP-expressing cells with erythroid characteristics.
- To establish optimal erythroid differentiation conditions for HEL and MEL cells.
- To characterize expression of β-globin from the vector-encoded *HBB* and concomitant suppression of transgenic GFP from shRNA-encoding MA821Q-derived vectors.

3.2.1 Introduction

RNAi-based silencing is a powerful tool for functional genetic screening in a number of organisms. Since the initial characterization of RNAi in the nematode *C. elegans*, great progress has been made towards the development of techniques for targeted gene knockdown. The expression of siRNAs from an inverted repeat sequence containing a hairpin allowed the development of intracellularly-expressed silencing triggers conferring stable and long-lasting silencing. These structures, termed shRNAs, and their ability to reduce the expression of target genes in a sequence-specific fashion have become critical tools in the characterisation of gene function and allow the development of therapeutic applications for a wide range of disease genes.

The overriding objective of this study was the functional characterization of site and context of shRNA expression from *HBB* intron 2 of the MA821Q vector backbone. For quick and low-cost assessment, we designed shRNAs targeting GFP, a bright fluorophore with a half-life of about 26 hours and stable expression (Ex_{max} at 395 nm, Em_{max} at 510 nm [325]). The RNAi machinery uses the shRNA technology to cleave across homologous sites on the GFP transcript, reduce protein synthesis and levels of fluorescence. For GFP as a target, these changes can be assessed by flow-cytometric (or microscopic) measurement of GFP fluorescence, which

represents a simple and robust approach to quickly identify an effective silencing trigger [326]. The shRNA stem sequence is a critical determinant of potent target knockdown. However, the structural features affecting RNAi efficiency and therefore required to accurately predict shRNA functionality are not, as yet, clearly defined. We therefore employed an already functionally validated GFP target sequence [302] for incorporation into our shRNA expression vectors. In the following we functionally characterise the corresponding GFP-specific shRNA-coding vectors in GFP-transgenic mammalian cell-based systems.

3.2.2 GFP knockdown in BM cells of GFP transgenic mice

The UBI-GFP/HBB$^{th3/+}$ transgenic mouse line has a β-thalassaemia *intermedia* phenotype but carries a ubiquitously expressed GFP transgene [327] (Jackson Laboratory Stock number 2683) and can be used for BMT studies with LV-genetically-corrected cells to assess phenotypic correction of the disease. The original UBI-GFP mouse model was generated directly in a C57BL/6J strain, modified to express GFP under control of the human ubiquitin C promoter [328] (Jackson Laboratory Stock number 4353). These animals were crossed to generate the UBI-GFP/HBB$^{th3/+}$ mouse, which presented an ideal model system for characterizing the GFP-shRNA expression vector since no other suitable cell culture-based system was available at the time of commencement of this project.

BM lineage-negative (Lin$^-$) stem-progenitor cells from inbred UBI-GFP/HBB$^{th3/+}$ mice were harvested [329], transduced with the MA821Qsh-G LV and cultured in methylcellulose-based media for two weeks to allow their development into discrete colonies. However, analysis for shRNA-mediated GFP knockdown in individual colonies by flow cytometry was uninformative, since the number of cells per colony was too small for signal detection above instrument noise, exacerbated by technical problems with the machine at the time (data not shown). This technical limitation prompted the creation of immortal GFP-transgenic cell lines, which would provide potentially unlimited cell numbers.

3.2.3 Assessment of GFP-shRNA activity in erythroid-induced GFP-HEL cell cultures

In order to ensure an abundance of material for analysis, the human erythroleukemia (HEL) cell line was transduced with an LV to express GFP under the control of the constitutive PGK promoter (**section 2.8.2**), and individual expanded clones were isolated by two rounds of limiting dilution. The HEL cells were selected as they share characteristics with primary erythroid progenitors, namely, their human origin, differentiation along the erythroid lineage, and globin synthesis. The GFP-expressing HEL (GFP-HEL) cell line was used as a reporter system for large-scale analyses and long-term assessment of vector functionality.

3.2.3.1 Identification of effective inducers of erythroid differentiation

The MA821Q vector harbours the *HBB* transgene under the transcriptional control of the late erythroid *HBB* promoter and a reduced βLCR, and confers temporal control of expression during erythroid development. Optimal inducing conditions of erythroid differentiation in GFP-HEL cells were investigated by adding different bioactive reagents to the cultures, followed by benzidine staining for haeme synthesis, and, in some cases, by erythroid lineage-specific marker expression. Viability was investigated by trypan blue exclusion assay. This study compared four substances, namely, haemin (30 μM, 50 μM, and 100 μM [250, 330]), DMSO (0.5% [330]), sodium butyrate (NaButyrate; 0.5 mM [330, 331]) and EPO (10 U/mL [332]) at concentrations previously described to promote effective erythroid differentiation in comparison to control (uninduced) GFP-HEL cultures. In our hands, haemin was the only substance showing erythroid inducing activity ($F(6,11)=40.46$, $p<0.0001$), while for all substances cellular toxicity was low and comparable ($F(6,11)=1.732$, $p=0.2033$) (**Figure 3.2.1A**). Haemin effectively induced erythroid differentiation at 30 μM, 50 μM and 100 μM, producing 72.0±1.5%, 78.8±10.2% and 90.2±13.9% of haemoglobinized cells, respectively. As the percentage of Hb-producing cells and toxicity did not differ significantly between the tested haemin concentrations, we suggested that any of these was appropriate for inducing erythroid differentiation. We chose haemin at 50 μM concentration as a standard erythroid induction regimen since the majority of published work on erythroid induction was performed using this concentration, which would

allow direct comparison with work by others. However, we noticed that haemin induction did not result in cell pellets with a distinct red colouration in spite of an almost 80% of benzidine positivity, which might be caused by non-specific staining as a result of cellular uptake of exogenous haemin [333]. We therefore further investigated the level of Hb synthesis by using Western blotting on whole cell extracts. Our findings agree with published data [334], showing increased HBG synthesis with 50 µM haemin treatment (**Figure 3.2.1B**). Surprisingly, and in contrast to naïve HEL cultures, detectable HBG protein in control GFP-HEL cultures pointed towards the possibility that transgenesis alone could affect *HBG* expression. In fact, substances other than haemin exhibited a similar baseline *HBG* expression to control GFP-HEL cultures. Nevertheless, haemin clearly augmented HBG synthesis as indicated by a strong Western blot signal, which are in agreement with the findings deduced with the benzidine assay and identifying haemin as a potent erythroid inducer of GFP-HEL cells.

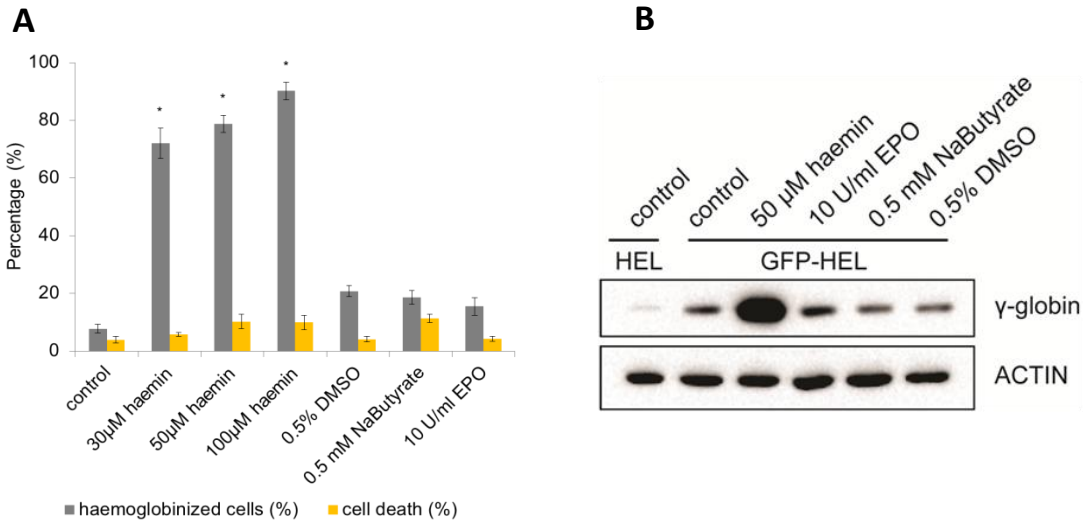


Figure 3.2.1. Erythroid induction of GFP-HEL cells. A) Values represent the percentage (%) of haemoglobinized cells (grey bars) and cell death (yellow bars) in cultures of GFP-HEL cells treated with test inducers of erythroid differentiation as indicated and compared to uninduced control cultures. Histograms represent means ± SD from 2–4 experiments in which each condition was repeated in triplicate. Statistical analysis was performed by one-way ANOVA. Statistically different values (vs. control) are annotated (*, $p<0.05$). **B)** Whole cell extracts of uninduced HEL cells and of GFP-HEL cells that were uninduced and or treated with different inducer of erythroid differentiation (as in **A**) analysed by Western blotting and screened for HBG (γ-globin) chains. Actin served as a loading control.

Furthermore, the progression of erythroid differentiation in GFP-HEL cells treated with different test inducer substances was assessed by flow-cytometric detection of erythroid-specific surface markers glycophorin A (GpA) and transferrin receptor (CD71). The analysis described here uses specific antibodies coupled to distinct fluorochromes, which are used concurrently to assess cell-surface marker co-expression within the same sample. Compensation analysis was performed to remove unwanted fluorophore signal from secondary detection channels. Assuming that HEL cells recapitulate surface-marker expression of erythroid progenitor cells during differentiation, we would expect high levels of CD71 but no GpA in control cultures, followed by a change towards a moderate expression for both CD71 and GpA upon erythroid differentiation [335]. Instead, we detected high levels of GpA expression in control cultures in two independent experiments, which between them exhibited a significant 1.6-fold difference in mean fluorescence intensity (MFI; 516.2±29.6 vs. 852.5±70.7, n=3) (**Figure 3.2.2A**). In order to explain the detected differences in GpA expression in cultures derived from the same GFP-HEL clone at different time-points, we proposed that phenotypic changes would occur during prolonged continuous culture, e.g. as stress erythropoiesis when culture conditions become suboptimal before a medium change. In comparison to control GFP-HEL cells, we detected an MFI-GpA reduction within the range of 45%–60% following treatment with haemin. In a 2nd experiment, the 50 µM haemin represented a positive control for erythroid differentiation and was used in parallel to other test compounds, namely DMSO (0.5%), NaButyrate (0.5 mM) and EPO (10 U/mL). In comparison to control GFP-HEL cells, we detected an MFI-GpA reduction for all treatments. DMSO led to a non-significant reduction of 11.5%, whereas EPO and NaButyrate led to a significant reduction of 25% and 59%, respectively (**Figure 3.2.2A**). Data analysis was performed by one-way ANOVA, and results for the two experiments reported as follows; 1st experiment: ($F(3,8)=119.8$, $p<0.0001$, white bars), 2nd experiment: ($F(4,10)=30.73$, $p<0.0001$, grey bars) (**Figure 3.2.2A**). Further analyses on CD71 erythroid-specific surface-marker expression exhibited high level of CD71 expression in control GFP-HEL cultures (MFI; 236±9.4, n=3), which decreased with increasing concentrations of haemin. Treatment with 30 µM, 50 µM and 100 µM haemin resulted in a significant reduction in MFI-CD71 by 13.3%, 29.2% and 45.2%, respectively (**Figure 3.2.2B**). In the second experiment, both 50 µM haemin and DMSO resulted in a significant reduction by 45% and 23.2%, respectively, whereas EPO and

NaButyrate resulted in a non-significant reduction in the range of 4%–6% (**Figure 3.2.2B**). These findings suggested that the CD71 expression profile of GFP-HEL cells transitioning from a control into an erythroid environment is not a suitable marker for erythroid differentiation, since a similar profile was also detected for non-erythroid inducers such as DMSO. Data analysis was performed by one-way ANOVA and results for the two experiments reported as follows; 1st experiment: ($F(3,8)=105.5$, $p<0.0001$, white bars), 2nd experiment: ($F(4,10)=29.05$, $p<0.0001$, grey bars) (**Figure 3.2.2B**). In summary, it would appear that the quantitative analysis of erythroid-specific surface marker expression by flow cytometry as a complement to benzidine assays is not informative for identifying potent erythroid inducers in HEL cells.

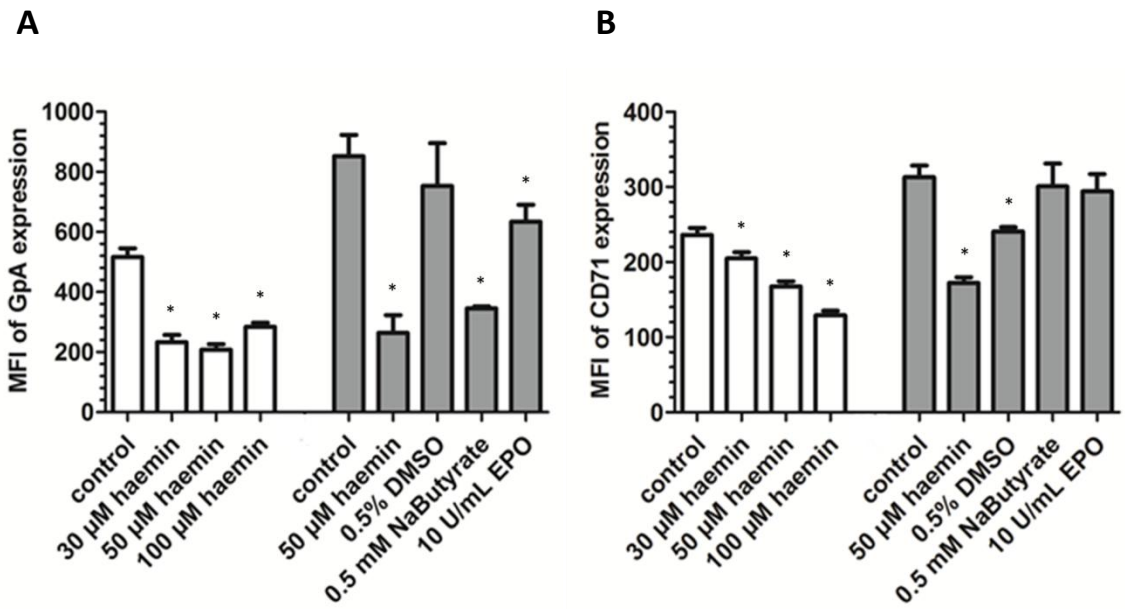


Figure 3.2.2. Flow cytometry analysis of GpA and CD71 erythroid-specific surface-marker expression in GFP-HEL cells. Values represent mean fluorescence intensity (MFI) for **A)** glycophorin A (GpA) and **B)** transferrin receptor (CD71). GFP-HEL cultures were treated with biological substances as indicated and compared to uninduced controls. Histograms represent means ± SD, and distinctly coloured bars represent data of different experiments. Each treatment was repeated in triplicate. Statistical analysis was performed by one-way ANOVA. Statistically different values (vs. control) are annotated (*, $p<0.05$).

3.2.3.2 The effect of erythroid induction on GFP-HEL cell phenotype

Flow cytometry analysis of GFP-HEL cells separated the population into two distinct clusters on a forward scatter (FSC)/side scatter (SSC) plot, despite HEL cells being an established line (**Figure 3.2.3A**). FSC correlates with cell size, and SSC with the density (cellular granularity and membrane size) of cells. In order to measure the distribution of the two cell subsets within the larger parent population quantitatively, we first eliminated unwanted events and defined populations of interest by placing scatter gates. The number of events detected in each region was measured on a histogram of count vs. FSC, and the proportion of cells in each gated population was analysed in comparison to the total number of cells in the parent population. The dominant cluster in naïve HEL cells displayed a SSC^{low}/FSC^{high} profile, constituting approximately 90% of the total population. In control GFP-HEL cultures, we detected a shift (without statistical significance) of approximately 20% of the SSC^{low}/FSC^{high} cells toward a SSC^{high}/FSC^{low} profile, suggesting transgenesis being as a contributing factor to changes observed in cellular phenotype (**Figure 3.2.3A**).

Next, we evaluated the distribution of both clusters following addition of bioactive reagents. Data analysis was performed for three independent experiments and each treatment was repeated thrice (**Figure 3.2.3B–D**)**.** The interactive effect between test compounds and cell phenotype was highly significant in all three experiments suggesting that both variables influence the dynamics of cell distribution and reported as follows: **Figure 3.2.3B**: ($F(3,16)=262.02$, $p<0.0001$), **Figure 3.2.3C**: ($F(4,20)=570.42$, $p<0.0001$); **Figure 3.2.3D**: ($F(4,20)=103.94$, $p<0.0001$). DMSO and EPO exhibited similar distribution patterns to control GFP-HEL cultures, whereas significant changes were detected in the presence of haemin and NaButyrate. In the 1st experiment, haemin shifted the distribution of cells at any concentration (30 µM, 50 µM, and 100 µM) with about 85% of cells acquiring a SSC^{high}/FSC^{low} profile while only 15% retained the control SSC^{low}/FSC^{high} profile (**Figure 3.2.3B**). Similar findings were detected with 50 µM hemin in the 2nd and 3rd experiments in which cells constituting the SSC^{high}/FSC^{low} cluster represented approximately 97% of the parental population (**Figure 3.2.3C**–**D**). As we found haemin to be a potent erythroid inducer in our HEL system, we assumed that observed phenotypic deviations from the parental profile were in consequence to changes occurring within cells undergoing

erythroid differentiation. However, a similar distribution pattern was detected in the presence of NaButyrate, which did not induce erythroid differentiation in the present HEL cultures. In the 2nd experiment, NaButyrate induced a change towards an almost equal cell distribution between the two clusters (40.4% of SSC^{high}/FSC^{low} vs. 59.1% of SSC^{low}/FSC^{high}) (**Figure 3.2.3C**) whereas, in the 3rd experiment, it produced a profound shift in distribution with about 95% of cells residing the SSC^{high}/FSC^{low} cluster (**Figure 3.2.3D**). Although the size of the effect was quite diverse between the two experiments, nonetheless, the effect on cell distribution was both significant and concurrent. It would appear that haemin exerts its effect on cellular phenotype through a process that is partially independent of terminal erythroid differentiation.

Our results are in accord with a previous study, which reported the presence of two HEL sub-populations distinct in their expression of the *c-kit* receptor (KR; SCF receptor, also known as CD117, the gene product of the c-kit proto-oncogene). The authors demonstrated that KR expression would change with the direction of cellular differentiation towards any of the erythroid, myeloid or megakaryocyte lineages depending on culture conditions; for example, highly-expressing KR cells were associated with a synchronous expression of the erythroid surface marker GpA, whereas megakaryocytic-induced differentiation resulted in decreased KR expression and concomitant upregulation of the megakaryocytic CD41b marker. They also reported spontaneous and reversible phenotypic changes in suspension cultures which were independent of clonal selection [336]. Based on these observations, we investigated KR expression in cultures of HEL cells treated with 50 μM haemin and compared the KR expression profiles of the two sub-populations, the prevalence of which in culture was previously shown to change with erythroid induction by haemin. Our results showed a dominant SSC^{high}/FSC^{low} sub-population in both control and haemin-treated cultures, constituting >98% of the total cell population. The basic KR expression in control HEL cells was 67.6±2.5 MFI-KR and was increased by 85.6% following haemin induction (MFI: 125.6±6.1) ($F(1,8)=443.92$, $p<0.0001$) (**Figure 3.2.4**). This finding was in agreement with the published observation of a higher KR expression in erythroid-differentiated cells, thus providing additional evidence of haemin having an erythroid-inducing activity on HEL cells. However, in the same paper, the authors did not relate KR expression with a distinct FSC vs. SSC phenotypic profile, for which our analyses

could likewise not find a correlation. We then hypothesised that, at the time of the experimental set-up, the HEL cells being of low passage (#2) were homogeneous in their FSC vs. SSC profile, since, as previously suggested, spontaneous phenotypic changes would occur with the propagation of cultures. The term passage number refers to the number of times cells have been removed from their culture plate and undergone a sub-culture after thawing. Follow up on this hypothesis, we induced HEL cells of a higher passage number (#7) with 50 µM haemin for 4 days, before analysis for c-kit expression by flow cytometry. In comparison to the control sample, we detected an approximately 1415% increase in the number of cells with a SSC^{high}/FSC^{low} profile (0.9±0.3% vs. 13.3±0.3%) and a concomitant loss of about 13% of the SSC^{low}/FSC^{high} cells (99.1±9.2% vs. 86.7±9.2%) following induction with 50 µM haemin. Analysis showed a significant interactive effect between haemin induction and cell phenotype, both contributing to the distribution pattern of cells on the FSC vs. SSC plot (**Figure 3.2.4**; $F(1,8)=83.8$, $p<0.0001$). Similarly to our previous findings, we detected an increase in KR expression in HEL cultures treated with 50 µM haemin, which resulted in 29.1% and 63.4% for the SSC^{high}/FSC^{low} and SSC^{low}/FSC^{high} sub-populations, respectively, and at 53.2% for the total cell population ($F(2,12)=78.96$, $p<0.0001$). These findings confirmed erythroid differentiation of cells using a 4-day induction with 50 µM haemin via up-regulation of the erythroid c-kit marker. Once again, the effect of haemin on cell phenotype appeared to be unrelated to the mechanism driving erythroid differentiation.

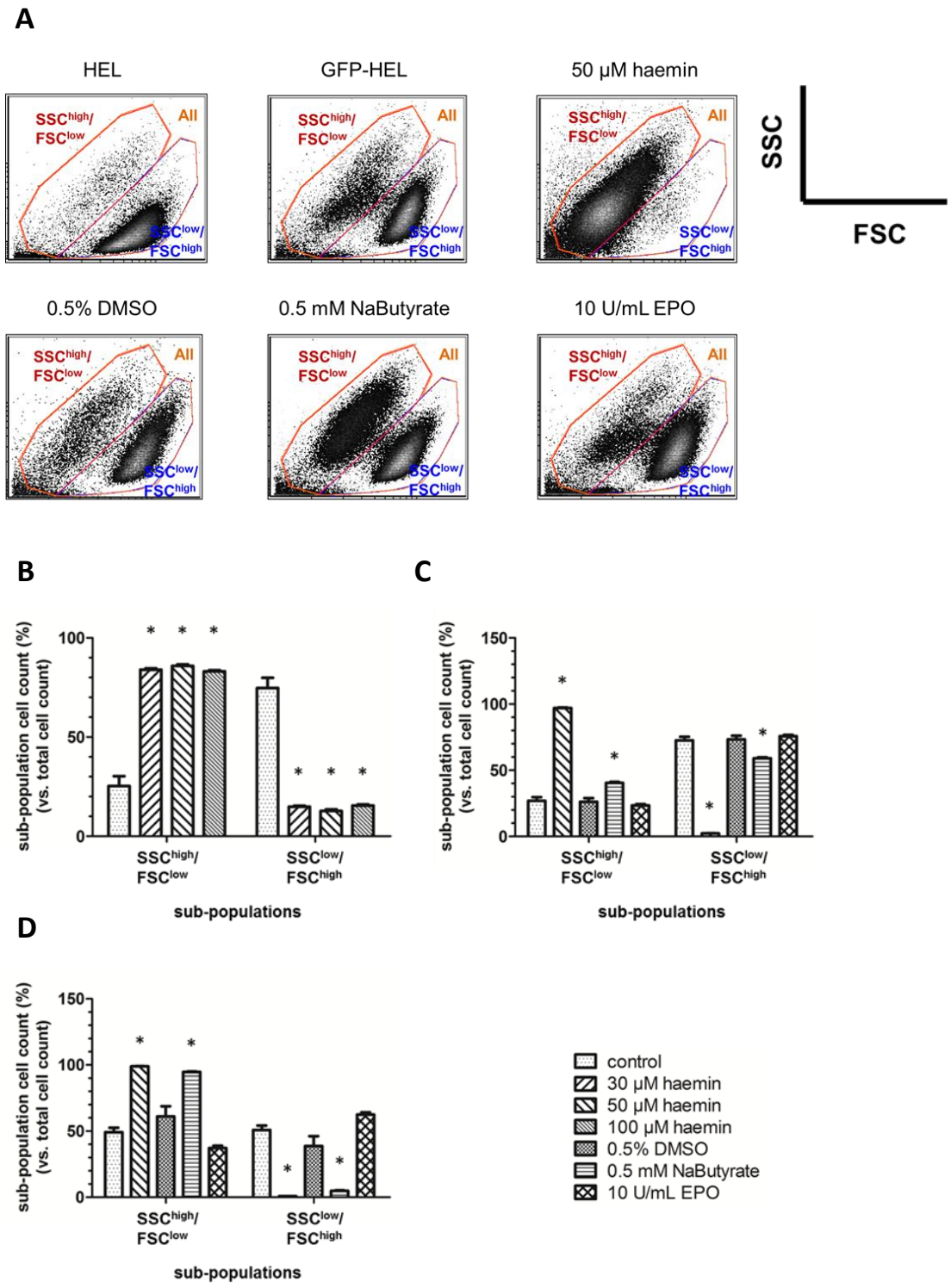


Figure 3.2.3. The effect of test inducer compounds on GFP-HEL cell phenotype. GFP-HEL cultures were treated with biological substances as indicated and compared to uninduced controls. **A)** FSC vs. SSC plots of gated live cells in the total cell population (All cells) and each sub-population (SSC/FSC) separately. **B–D)** Histograms represent mean ± SD, and each plot represents data from one experiment. Each treatment was performed in triplicate. Values represent the percentage (%) of cells in each sub-population in comparison to the total cell count. Statistical analysis was performed by two-way ANOVA. Statistically different values (vs. control) are annotated (*, $p<0.05$).

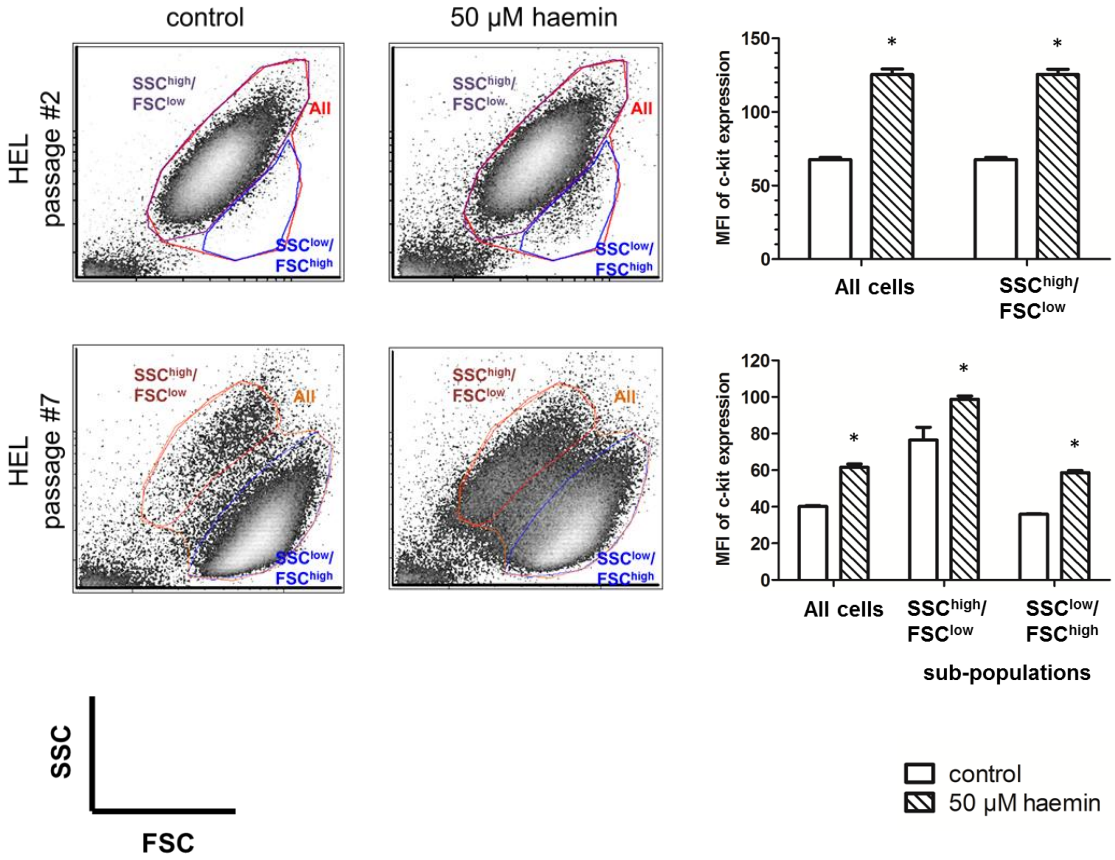


Figure 3.2.4. Flow cytometry analysis of c-kit erythroid-specific surface marker expression in HEL cells. FSC vs. SSC plots of gated live cell populations in HEL cell cultures of passage 2 and 7. Cultures were induced with 50 µM haemin for erythroid differentiation. Histograms represent mean ± SD for one experiment of triplicates. Values represent the mean fluorescence intensity (MFI) of c-kit expression in the total cell population (All cells) and each sub-population (SSC/FSC) separately. Statistical analysis was performed by two-way ANOVA. Statistical different values (vs. controls) are annotated (*, $p<0.05$).

3.2.3.3 The effect of erythroid induction on GFP expression in GFP-HEL cells

GFP expression in transgenic GFP-HEL cultures was followed by live fluorescence microscopy and quantitatively measured by flow cytometry. Routine monitoring by microscopy demonstrated a weaker GFP signal in cultures treated with haemin compared to controls (data not shown). Optical findings were further confirmed by flow cytometry. In the 1st experiment, erythroid induction with haemin significantly reduced GFP fluorescence at 30 μM, 50 μM and 100 μM as compared to control GFP-HEL cultures (**Figure 3.2.5A**). In the total cell population, the reduction in MFI-GFP was up to 98% at any haemin concentration. In the two sub-populations, the reduction in MFI-GFP became larger with increasing concentrations of haemin. At 30 μM the decrease was 74% and 49% in the SSC^{high}/FSC^{low} and SSC^{low}/FSC^{high} sub-populations, respectively. Increasing the haemin concentration to 50 μM heightened reduction by approximately 83% in both sub-populations, while a further increase to 100 μM only marginally added to the MFI-GFP loss which reached up to 88% in both sub-populations ($F(6,24)=38.58$, $p<0.0001$) (**Figure 3.2.5A**). Evaluation of MFI-GFP in control GFP-HEL cultures demonstrated significant difference in the baseline fluorescence between the two sub-populations, with the SSC^{high}/FSC^{low} cluster exhibiting 92.5% lower fluorescence intensity compared to the SSC^{low}/FSC^{high} cluster ($p<0.0001$). Similarly, in cultures treated with haemin, the difference in MFI-GFP between the two sub-populations was retained in the range of 92%–96%, indicating that erythroid induction did not mediate changes in GFP expression in conjugation with an effect on cell redistribution. Importantly, overall we noted that treatment with haemin had a detrimental effect on GFP expression.

Homogenous GFP-HEL cell clones were generated by vector transduction and permanent integration of the *GFP* transgene in the cellular genome, such that GFP expression was stable, continuous (controlled by the PGK promoter) and heritable to daughter cells. Therefore we elaborated on our investigation on the phenomenon of *GFP* suppression with propagation of cultures and erythroid differentiation of cells.

First, we investigated the effect of culture propagation on *GFP* expression by comparing MFI-GFP in the control GFP-HEL cultures for three independent experiments, each cultured and

analysed at different passages. Analysis was performed by one-way ANOVA and results demonstrated that the extent of cell sub-culturing was not a contributing factor to the MFI-GFP reduction detected in cultures treated with haemin ($F(2,6)=0.5653$, $p=0.5958$) (data not shown).

Next, we investigated the effect of substances (DMSO, NaButyrate, and EPO), which did not induce erythroid differentiation in the present HEL cultures, on GFP expression. In the 2nd experiment, cultures treated with 50 μM haemin were included as positive controls of erythroid differentiation, whereas non-treated cultures were included as controls for basal levels of *GFP* transgene inactivation. Photographs captured by fluorescence microscopy exhibited a weak GFP signal in cultures treated with haemin compared to any other condition (**Figure 3.2.5C**). The MFI-GFP of these cultures was quantified by flow cytometry (**Figure 3.2.5B**). The interactive effect between biological substance and cell phenotype was highly significant suggesting that both variables contributed to changes in MFI-GFP ($F(8,30)=13.50$, $p<0.0001$). Following haemin induction, we detected a significant reduction in the MFI-GFP of the total cell population at ~20%, which validated the fluorescence imaging data. In the two sub-populations the reduction in MFI-GFP was 15% and 3% for the SSC^{low}/FSC^{high} and dominant SSC^{high}/FSC^{low} sub-populations, respectively (**Figure 3.2.5B**). Moreover, EPO did not affect GFP expression, resulting in a non-significant 1.5% loss, whereas both NaButyrate and DMSO led to a significant increase in MFI-GFP by 24.6% and 44.8%, respectively, in the total cell population (**Figure 3.2.5B**).

As there was a weak association between these two former experiments, a 3rd experiment (**Figure 3.2.5D**) was set-up in a similar fashion to the 2nd experiment (**Figure 3.2.5B**). The results we obtained showed a significant loss of MFI-GFP following treatment with haemin by approximately 98% in the total cell population as well as in the two sub-populations ($F(8,30)=88.97$, $p<0.0001$). Also, the SSC^{high}/FSC^{low} sub-population exhibited ~60% lower MFI-GFP compared to the SSC^{low}/FSC^{high} sub-population. Strikingly, and in contrast to the 2nd experiment, we detected a significant reduction in MFI-GFP following treatment with NaButyrate and DMSO by 95% and 35%, respectively, in the total cell population. Similarly though, EPO did not affect GFP expression (**Figure 3.2.5D**).

As the 2nd and 3rd experiments produced different findings for the same substances that appeared incapable of inducing erythroid differentiation, we next analysed GFP expression at the protein level using Western blot analysis (**Figure 3.2.5E**). Trials to determine the optimum conditions for GFP detection on Western blots showed that loading a protein content equivalent to 0.5–1x10^4 lysed cells was sufficient to detect differences in band intensities indicating the relative abundance of GFP protein (data not shown). However, as a larger amount of cell lysate was required for actin detection, using the same membrane for GFP staining would have resulted in saturation of the GFP signal. Therefore we prepared two gels for detection of target and control proteins separately. The number of lysed cells appropriate for detection was determined and corresponding volumes loaded for the two assays. To our surprise, we detected higher amounts of GFP protein in haemin- and NaButyrate-treated cultures compared to control GFP-HEL cultures (**Figure 3.2.5E**), an observation that contradicted our findings by flow cytometry (**Figure 3.2.5D**). The absence of a logical explanation for why the positive control for GFP expression would display lower gene product in spite of an equal actin loading across samples, raised concerns over the sensitivity of the Western blot assay to detect changes in GFP expression. Bearing these observations in mind, we performed a trial experiment in which protein lysates from MEL cells, GFP-MEL cells, and a mixture of the two (1:1, v/v) were analysed on the same Western blot (**Figure3.2.5F**). This showed that naïve MEL cultures did not express the GFP transgene, whereas the mixed sample exhibited lower GFP protein product compared to the GFP-MEL sample, thus affirming the sensitivity of the Western blot assay to detect changes in GFP expression (**Figure 3.2.5F**). Despite this finding confirming the general reliability of our immunoblots and with fluorescence imaging and flow cytometry being the more direct approaches and producing similar data, we chose to prioritise flow cytometry for the production of protein data in future experiments.

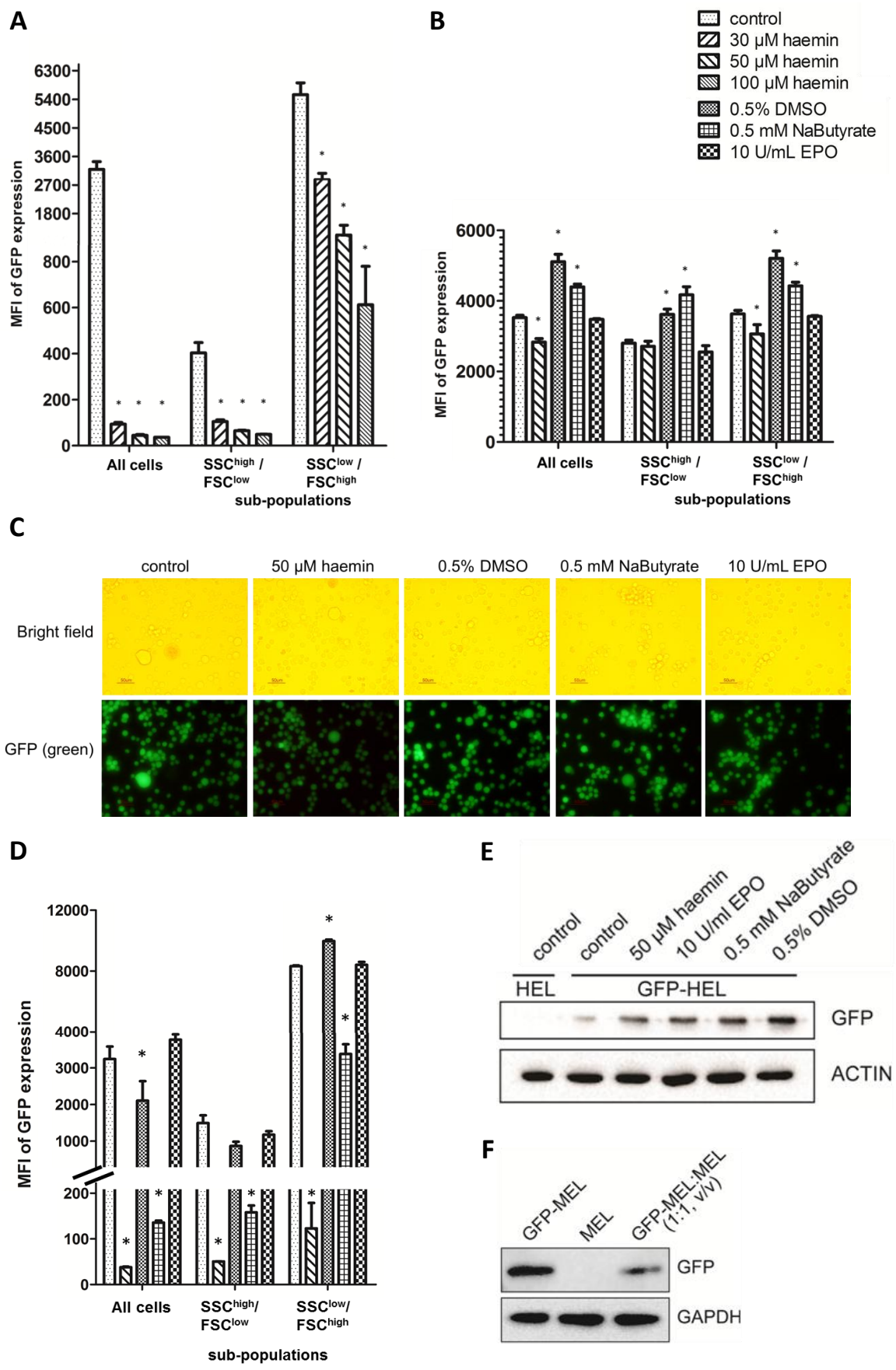

A
MFI of GFP expression
6300
5400
4500
3600
2700
1800
800
600
400
200
0
All cells
SSC^high / FSC^low
SSC^low / FSC^high
sub-populations
B
control
30 μM haemin
50 μM haemin
100 μM haemin
0.5% DMSO
0.5 mM NaButyrate
10 U/mL EPO
MFI of GFP expression
6000
4000
2000
0
All cells
SSC^high / FSC^low
SSC^low / FSC^high
sub-populations
C
control
50 μM haemin
0.5% DMSO
0.5 mM NaButyrate
10 U/mL EPO
Bright field
GFP (green)
D
MFI of GFP expression
12000
8000
4000
3000
2000
1000
200
100
0
All cells
SSC^high/ FSC^low
SSC^low/ FSC^high
sub-populations
E
control
control
50 μM haemin
10 U/ml EPO
0.5 mM NaButyrate
0.5% DMSO
HEL
GFP-HEL
GFP
ACTIN
F
GFP-MEL
MEL
GFP-MEL-MEL (1:1, v/v)
GFP
GAPDH

Figure 3.2.5. Analysis of GFP expression in GFP-HEL cells treated with bioactive reagents. A, B, D) Histograms represent mean ± SD for three independent experiments. Each treatment was performed in triplicate. Values represent mean fluorescence intensity (MFI) for GFP. Statistical analysis was performed by two-way ANOVA. Statistical different values (vs. controls) are annotated (*, $p<0.05$). **C)** Fluorescence microscopy of GFP-HEL cells treated with bioactive reagents as indicated. Images are captured by light microscopy (top row) and fluorescence microscopy (bottom row). Amplification factor of the microscope was 20x. Scale bars = 50 µM. **E)** Whole cell extracts of HEL and GFP-HEL cells analysed by Western blotting and tested for GFP. Actin served as a control for equal loading. **F)** Whole cell extracts of MEL and GFP-MEL cells analysed by Western blotting and tested for GFP. GAPDH served as a control for equal loading.

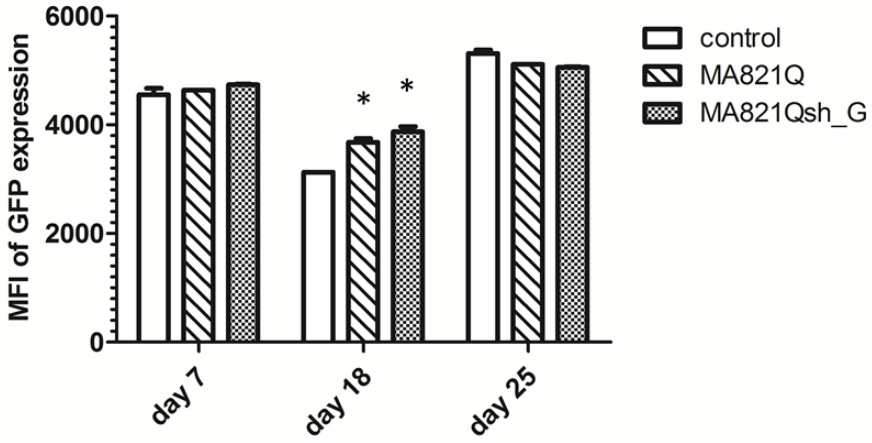
3.2.4 Assessment of GFP-shRNA activity in naive GFP-HEL cells

In the previous sections we demonstrated that haemin is a potent erythroid inducer of GFP-HEL cells. In addition, we showed that haemin contributes to inactivation of the *GFP* transgene. As the profound reduction in MFI-GFP would possibly obscure siRNA silencing activity, we next transduced GFP-HEL cells with our GFP-specific shRNA vector in cultures in the absence of erythroid inducers and any associated toxicities, followed by GFP knockdown analysis. In contrast to endogenous *HBB*, the vector-encoded gene, lacking repressing control elements, would be expressed at least at basal levels in naïve HEL cells.

Our HIV-based system is pseudotyped with the vesicular stomatitis virus glycoprotein (VSV-G), which has binding affinity to a ubiquitous cell surface receptor, and thus ability to transduce almost all cell types [337]. The murine erythroleukemia (MEL) cells line has been used extensively in screening studies with erythroid-specific vectors using an MOI close to 1 to achieve single proviral integrations per cell [338]. In order to ensure stable transmission and elevated transgene expression, which would enable detection of even sub-optimal knockdown, GFP-HEL cultures were transduced at a MOI of 5. Transductions were performed with the i) MA821Q vector as control for *HBB* expression, and ii) MA821Qsh-G vector for RNAi activation and *GFP* knockdown. Both vectors encode the mutant *HBB* (HBB^{T87Q}) to facilitate discrimination between transgenic and endogenous HBB. Un-transduced naïve (control) GFP-HEL cells were included as negative controls for baseline gene expression. Transduced GFP-HEL cells were maintained in culture and GFP expression followed for up to 25 days post-transduction in order to investigate the stability of vector-derived *HBB* expression and the extent of RNAi silencing.

Transductions were performed in sets of triplicates and outlier removal was included in the analysis to reduce error variance and enhance the power of statistical analysis. GFP expression was measured by flow cytometry and analysed on days 7, 18 and 25 post-transduction (**Figure 3.2.6**). The correlation between duration of culture and vector type was highly significant, suggesting that both variables contribute to changes in GFP fluorescence ($F(4,10)=10.68$, $p=0.0012$). In comparison to control GFP-HEL cells, the MA821Qsh-G-transduced cultures exhibited an MFI-GFP increase of 4.1% at day 7, which had increased to 24% at day 18. End-

point analysis on day 25, which measures outcomes after continuous selection of the most rapidly dividing cells in the pool, exhibited an MFI-GFP decrease by 4.8%. We detected oscillation in GFP expression over the course of the culture period, decreasing by 18.2% from day 7 to day 18 and increasing by 30.5% from day 18 to day 25. A similar MFI-GFP profile was detected for the MA821Q-transduced cultures (**Figure 3.2.6**), indicating that changes in MFI-GFP were unlikely to be associated with shRNA activity. In addition, because similar MFI-GFP oscillation was detectable in control GFP-HEL cultures, we speculated that changes in MFI-GFP could develop over a continuous culture rather than associated with the process of transduction. Importantly, transduction with MA821Qsh-G did not produce significant reduction in MFI-GFP, indicating that the RNAi strategy was in all likelihood ineffective.



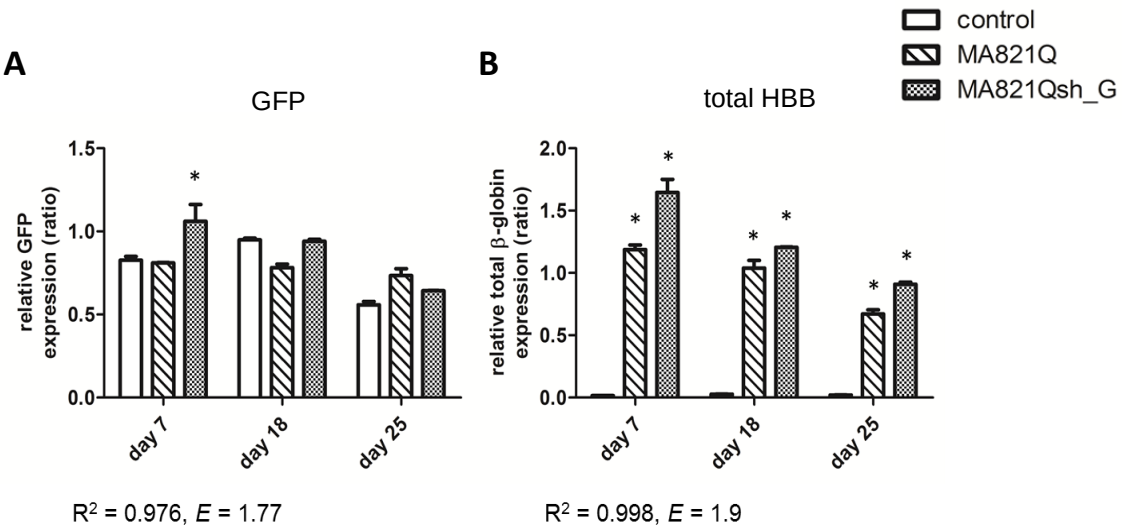
	MFI-GFP					
	Day 7		Day 18		Day 25	
	MFI	Δ% (fold change)	MFI	Δ% (fold change)	MFI	Δ% (fold change)
control	4551		3125		5310	
MA821Q	4636	1.89 (1.02)	3676	17.62 (1.18)	5111	-3.75 (0.96)
MA821Qsh_G	4737	4.11 (1.04)	3875	23.97 (1.24)	5056	-4.79 (0.95)

Figure 3.2.6. Flow cytometry analysis of GFP expression in GFP-HEL cells transduced with the MA821Q and MA821Qsh_G LVs at a MOI of 5. Cells were analysed for GFP mean fluorescence intensity (MFI-GFP) by flow cytometry on days 7, 18 and 25 post-transduction. Control for baseline expression were untransduced naïve GFP-HEL cells. The histogram represents mean ± SD for one experiment of triplicates. The table shows MFI-GFP, the percentage change (Δ%) and fold change in MFI-GFP between the transduced and the control cultures. Statistical analysis was performed by two-way ANOVA. Statistically different values (vs. control) are annotated (*, p<0.05).

Next, we performed analysis for GFP mRNA transcript levels by RT-qPCR. In order to detect the GFP transcript, primers were designed to overlap and amplify across the siRNA cleavage site on the target RNA. Independent studies show that 3' mRNA cleavage products are not degraded but instead accumulate within cells and act as templates for cDNA synthesis, thus leading to an underestimation of siRNA efficacy [339, 340]. In comparison to control GFP-HEL cells, day-7 analysis for the MA821Qsh-G transduced cultures showed elevated GFP expression by 28.35%, day-18 analysis revealed a decrease by 24%, whereas day 25 analysis showed an increase by 15.2% (**Figure 3.2.7A**). A similar pattern in GFP expression was also detected for the MA821Q-transduced cultures (**Figure 3.2.7A**)**.** Importantly, we did not detect significant reduction in GFP expression in cultures transduced with the MA821Qsh_G.

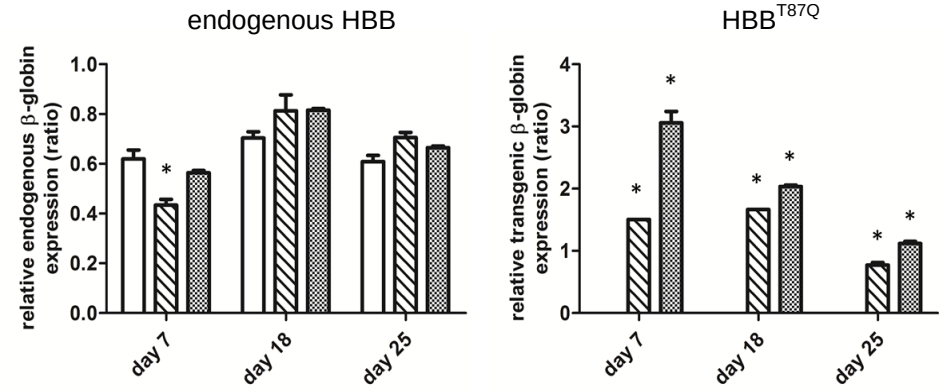
In order to evaluate that the lariat intron was properly spliced, we next measured the cellular HBB content at each time point. The primers were designed to amplify both endogenous and vector-derived HBB (total HBB), showing bulk accumulation of total HBB transcripts in both the MA821Q- and MA821Qsh_G-transduced cultures (**Figure 3.2.7B**). We detected a significant reduction in total HBB mRNA by 43.5% and 44.7% (p<0.0001) for MA821Q- and MA821Qsh_G-transduced cultures respectively, from day 7 to day 25. Following long-term propagation in culture, the transduced cell population appeared to become less heterogeneous for HBB^{T87Q} expression. During transduction, some cells would have received more than one vector copy or possibly none, contributing to cellular variation in HBB content. In addition, the expression of integrated provirus would be affected by the integration site and its surrounding chromatin, which could trigger epigenetic modifications leading to either inefficient transcription or even silencing of the HBB^{T87Q} transgene [341]. We speculated that HBB measured was mainly vector-derived, since many publications reported the absence of adult Hb in HEL cells [250, 330, 334] (exception for [342] showing HBB expression in particular sub-strains of HEL cells). As shown in **Figure 3.2.7B**, low levels of total HBB mRNA were also detected in control GFP-HEL cells, indicating that our HEL cell line was capable of HBB expression. This observation was confirmed using primers specific for the endogenous HBB mRNA transcript (**Figure 3.2.7C**). Importantly, we detected a significant increase in endogenous HBB mRNA by 87.6% (p<0.001) and 44.6% (p<0.01) for MA821Q- and MA821Qsh_G-transduced cultures respectively, from day

7 to day 18, which persisted over the remaining course of culture. Next, we measured the expression of transgenic *HBB* (**Figure 3.2.7D**). We did not detect product amplification in the control cultures, verifying that the primers used were specific for HBB^{T87Q}. Although HBB^{T87Q} mRNA was detected in both MA821Q- and MA821Qsh_G-transduced cultures, significantly higher amounts were produced with the MA821Qsh_G transduction, possibly due to a higher vector copy per cell (data not available). Finally, the HBB^{T87Q} expression profile was similar to the one detected for the total HBB, suggesting that the contribution of the endogenous *HBB* expression to the total HBB output was negligible.



R^2 = 0.976, *E* = 1.77

R^2 = 0.998, *E* = 1.9



R^2 = 0.937, *E* = 1.8

R^2 = 0.967, *E* = 1.89

Day 7	relative gene expression						
	GFP		total HBB		endogenous HBB		HBB^{T87Q}
	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio
control	0.83		0.02		0.62		0.00
MA821Q	0.81	-1.95 (0.98)	1.19	7073.67 (71.74)	0.43	2516.93 (26.17)	1.50
MA821Qsh_G	1.06	28.35 (1.28)	1.65	9840.44 (99.40)	0.56	3304.73 (34.05)	3.06

Day 18	relative gene expression						
	GFP		total HBB		endogenous HBB		HBB^{T87Q}
	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio
control	0.95		0.03		0.70		0.00
MA821Q	0.78	-17.65 (0.82)	1.04	3572.41 (36.72)	0.81	2776.22 (28.76)	1.66
MA821Qsh_G	0.94	-0.89 (0.99)	1.21	4165.68 (42.66)	0.82	2783.32 (28.83)	2.03

Day 25	relative gene expression						
	GFP		total HBB		endogenous HBB		HBB^{T87Q}
	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio
control	0.56		0.02		0.61		0.00
MA821Q	0.73	31.42 (1.31)	0.67	3221.62 (33.22)	0.71	3393.00 (34.93)	0.77
MA821Qsh_G	0.64	15.21 (1.15)	0.91	4403.91 (45.04)	0.66	3192.05 (32.92)	1.12

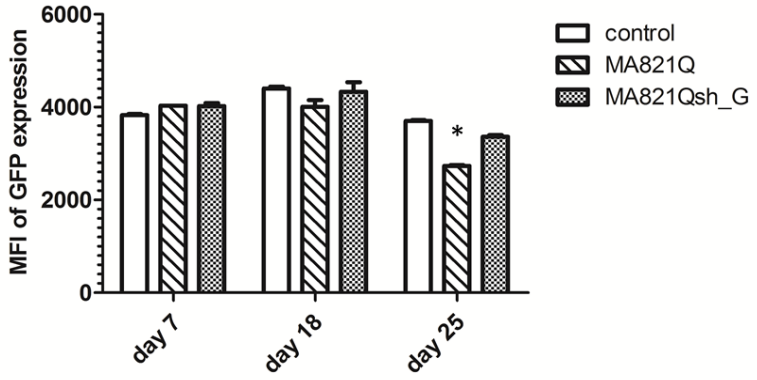
Figure 3.2.7. RT-qPCR analysis of mRNA expression in GFP-HEL cells transduced with the MA821Q and MA821Qsh_G LVs at a MOI of 5. Cells were analysed for **A)** GFP, **B)** total HBB (endogenous and vector-derived), **C)** endogenous HBB and **D)** vector-derived HBB (HBB^{T87Q}) mRNA levels by RT-qPCR on days 7, 18 and 25 post-transduction. Values are normalized by *GAPDH* expression. Histograms represent mean ± SD for one experiment of triplicates. Tables show relative gene expression, the percentage change (Δ%) and fold change in gene expression between the transduced and the control cultures. Statistical analysis was performed by two-way ANOVA. Statistically different values (vs. control) are annotated (*, $p<0.05$). E: PCR amplification efficiency; R^2: square of Pearson's correlation coefficient.

In case of reduced shRNA activity as a result of limited accessibility to target RNAs and/or inefficient processing compared to the mRNA rate of transcription, a weak target knockdown will fail to reduce the steady-state concentration of mRNA transcripts. Therefore we transduced naïve GFP-HEL cells with the MA821Q and MA821Qsh_G vectors at a higher MOI of 50. Transduced GFP-HEL cells were maintained in cultures and GFP expression was followed for up to 25 days post-transduction. Transductions were performed in triplicate sets and outliers removed in the analysis to reduce error variance and enhance accuracy of estimates. GFP expression was measured by flow cytometry and analysed on days 7, 18 and 25 post-transduction (**Figure 3.2.8**). In comparison to control GFP-HEL cells, day-7 analysis for the MA821Qsh-G-transduced cultures exhibited an MFI-GFP increase by 5.2%, whereas day-18 analysis revealed a small decrease by 1.6%. End-point analysis on day 25 exhibited an MFI-GFP decrease by 9.1%. We detected loss in MFI-GFP over the culture period, decreasing by 16.4% (p<0.001) from day 7 to day 25. A similar MFI-GFP profile was detected for the MA821Q-transduced cultures in which day-25 analysis exhibited significant MFI-GFP decrease by 26.1% (p<0.001) (**Figure 3.2.8**). Non-detectable GFP knockdown with MA821Qsh_G transduction using high MOI endorsed the possibility for a non-functional RNAi strategy.

We then measured the expression of mRNA encoding GFP and total HBB for days 18 and 25. Because of cytotoxicity and delayed cell growth subsequent to transduction with excess virus, reduced cell material at the beginning of cultures impeded day-7 analysis. In comparison to control GFP-HEL cells, day-18 and day-25 analyses for the MA821Qsh-G-transduced cultures showed elevated *GFP* expression by 114.04% (p<0.05) and 33.8%, respectively. Also, we detected loss in *GFP* expression over the culture period, decreasing by 45.3% (p<0.05) from day 18 to day 25 (**Figure 3.2.9A**). However, analysis at day 18 of the MA821Q-transduced cultures showed elevated *GFP* expression by 55.7%, whereas day-25 analysis showed decreased *GFP* expression by 19%. Similarly, we detected loss in *GFP* expression over the culture period, decreasing by 54.4% (p<0.05) from day 18 to day 25 (**Figure 3.2.9A**). Detectable changes in *GFP* expression with a vector system that did not harbour RNAi activity led us to conclude that GFP mRNA changes with the MA821Qsh_G-transduced cultures were not necessarily attributable to effective shRNA knockdown.

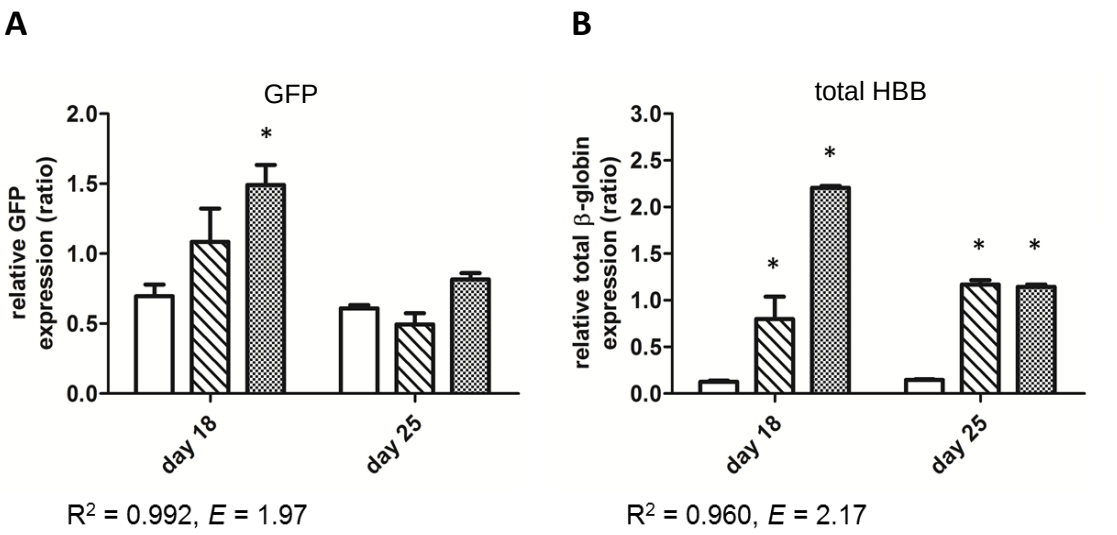
Furthermore, analysis for *HBB* expression revealed bulk accumulation of total HBB mRNA transcripts in both the MA821Q- and MA821Qsh_G-transduced cultures in comparison to the low total (endogenous) HBB expression in control cultures (**Figure 3.2.9B**). Transitioning from day 18 to day 25, the MA821Q-transduced cultures showed increased total HBB mRNA levels by 46.4%, whereas the MA821Qsh_G cultures showed significantly decreased amounts of total HBB mRNA by 48.1% (p<0.001). Assuming greater HBB^{T87Q} expression with increasing vector copies per cell, we showed that the decrease in total HBB expression in MA821Qsh_G transduced cultures was correlated with reduction in vector copies from ~50 (day 18) to ~15 (day 25). In MA821Q-transduced cultures, a smaller decrease in vector copies from ~31 (day 18) to ~24 (day 25) was insufficient to produce significant changes in total HBB expression (**Figure 3.2.9B**).

Studies in naïve GFP-HEL cell cultures identified proper processing of the HBB^{T87Q} mRNA from MA821Qsh_G, which, however, did not correlate with measurable GFP knockdown.



	MFI-GFP					
	Day 7		Day 18		Day 25	
	MFI	Δ% (fold change)	MFI	Δ% (fold change)	MFI	Δ% (fold change)
control	3826.3		4403.0		3700.7	
MA821Q	4030.1	5.32 (1.05)	4006.4	-9.01 (0.91)	2733.7	-26.13 (0.74)
MA821Qsh_G	4024.4	5.18 (1.05)	4335.0	-1.55 (0.98)	3362.8	-9.13 (0.91)

Figure 3.2.8. Flow cytometry analysis of GFP expression in GFP-HEL cells transduced with the MA821Q and MA821Qsh_G vectors at a MOI of 50. Cells were analysed for GFP mean fluorescence intensity (MFI-GFP) by flow cytometry on days 7, 18 and 25 post-transduction. The histogram represents mean ± SD for one experiment of triplicates. The table shows MFI-GFP, the percentage change (Δ%) and fold change in MFI-GFP between the transduced and the control cultures. Statistical analysis was performed by two-way ANOVA. Statistically different values (vs. control) are annotated (*, p<0.05).



Day 18	relative gene expression				VCN
	GFP		total HBB		
	ratio	Δ% (fold change)	ratio	Δ% (fold change)	
control	0.7		0.13		
MA821Q	1.1	55.67 (1.56)	0.80	526.34 (6.26)	31.24
MA821Qsh_G	1.5	114.04 (2.14)	2.21	1630.40 (17.30)	50.52

Day 25	relative gene expression				VCN
	GFP		total HBB		
	ratio	Δ% (fold change)	ratio	Δ% (fold change)	
control	0.61		0.15		
MA821Q	0.49	-18.96 (0.81)	1.17	700.31 (8.00)	23.67
MA821Qsh_G	0.82	33.76 (1.34)	1.14	684.23 (7.84)	15.91

Figure 3.2.9. RT-qPCR analysis of mRNA expression in GFP-HEL cells transduced with the MA821Q and MA821Qsh_G vectors at a MOI of 50. Cells were analysed for **A)** GFP and **B)** total HBB (endogenous and vector-derived) mRNA levels by RT-qPCR on days 18 and 25 post-transduction. Values are normalized by *GAPDH* expression. Histograms represent mean ± SD for one experiment of triplicate sets. Tables show relative gene expression, the percentage change (Δ%) and fold change in gene expression between the transduced and the control cultures, as well as the average cell VCN in bulk samples. Statistical analysis was performed by two-way ANOVA. Statistically different values (vs. control) are annotated (*, $p<0.05$). E: PCR amplification efficiency; R^2: square of Pearson's correlation coefficient.

3.2.5 Assessment of GFP-shRNA activity in erythroid-induced GFP-MEL cell cultures

In response to the unexpected setbacks in the HEL system, a murine erythroleukemia (MEL) cell line was employed instead. The MEL cell lines are derived from Friend virus-infected mice that results in transformation of erythroid precursor cells. The deregulated expression of the transcription factor PU.1 negatively affected the transcriptional activity of the erythroid regulator GATA1, giving rise to malignant erythroblasts and developmental arrest at the proerythroblast stage of differentiation [252, 343]. MEL cells treated with chemical agents such as DMSO and hexamethylene bisacetamide (HMBA) cause induced terminal erythroid differentiation that involves cessation of growth following an obligatory two rounds of mitosis. A characteristic feature of differentiated MEL cells is elevated adult *Hbb* expression, resulting in >10,000 copies of mRNA per cell within 3–4 days of induction, and absence of detectable embryonic (Hbbh1 and Hbey) globin equivalents [344, 345]. Furthermore, differentiated MEL cells harbouring stably integrated human *HBB*-locus transgenes showed co-expression of the murine and human adult HBB-like chains, suggesting that MEL cells retain transcriptional machinery analogous to the one operating in transgenic mice [330, 346]. These observations demonstrated that MEL cells have an erythroid differentiation programme similar to its normal equivalent, in spite of being transformed, and supported its use as a mammalian cell model for the study of transcriptional and post-transcriptional regulation of *HBB* expression.

3.2.5.1 Identification of an effective erythroid inducing regimen

We produced a βLCR-GFP-MEL cell line with inducible GFP expression in order to monitor differentiation and vector-derived expression. We used the LV MA1047 (provided by Dr. M. Antoniou, unpublished data), which contains the 0.7-kb coding sequence of the GFP reporter gene inserted between the first and third exons of an *HBB* backbone under control of the *HBB* promoter and HS2 and HS3 βLCR elements, akin to that of the GLOBE therapeutic LV. MEL cells were transduced, and individually expanded clones, isolated by one round of limiting dilution, were treated with different concentrations of DMSO to identify an effective erythroid-inducing regimen.

First, βLCR-GFP-MEL cells were cultured in 5%, 10% and 20% FBS-supplemented IMDM growth medium for 4 days, followed by analysis for spontaneous differentiation by benzidine staining and for GFP expression by flow cytometry. Each condition was performed in triplicate. Published work has demonstrated that MEL cells would exhibit spontaneous differentiation, albeit at low frequencies in the range of 1%–2% according to the MEL clone and culture conditions [347, 348]. In our study, we did not detect a significant difference in the percentage of Hb-producing cells at different culture conditions ($F(2,6)=4.205$, $p=0.0722$). Using 10% FBS, which represents a standard condition for general cell culture with common cell lines including MEL lines, we detected 2.3±1.3% Hb-producing cells, which was within the range of expected levels of spontaneous differentiation. Also, cultures of 5% FBS and 20% FBS produced a lower percentage of Hb-producing cells in the range of 1.3%–2.1% and 0.0%–1.1% respectively, (**Figure 3.2.10A**). In addition, cultures in the presence of 10% FBS produced the highest percentage of cell death at 1.6±0.7% and which was not significantly different to cell death observed for the other two conditions (0.0%–1.2%) ($F(2,6)=2.739$, $p=0.1429$) (**Figure 3.2.10A**). Furthermore, drawing on previous observations with GFP-HEL cells, we made the assumption that the reduced βLCR will retain transcription-enhancing activity to drive GFP expression under a sub-optimal erythroid environment. We detected GFP expression in any of the serum-containing cultures and reported a significantly lower MFI-GFP with 10% FBS ($F(2,6)=12.51$, $p=0.0072$). Post-hoc Tukey's test showed that cultures with 10% FBS had ~15.6% lower MFI-GFP than those with 5% FBS ($p<0.01$), while all other comparisons were not significant (**Figure 3.2.10B**). Having observed that the 10% FBS culture condition did not behave differently to a higher (20%) or lower (5%) FBS concentration in terms of spontaneous differentiation and cell death, while also exhibiting less promoter activity under undifferentiated conditions, we chose it as a standard condition for GFP-MEL cultures in subsequent experiments.

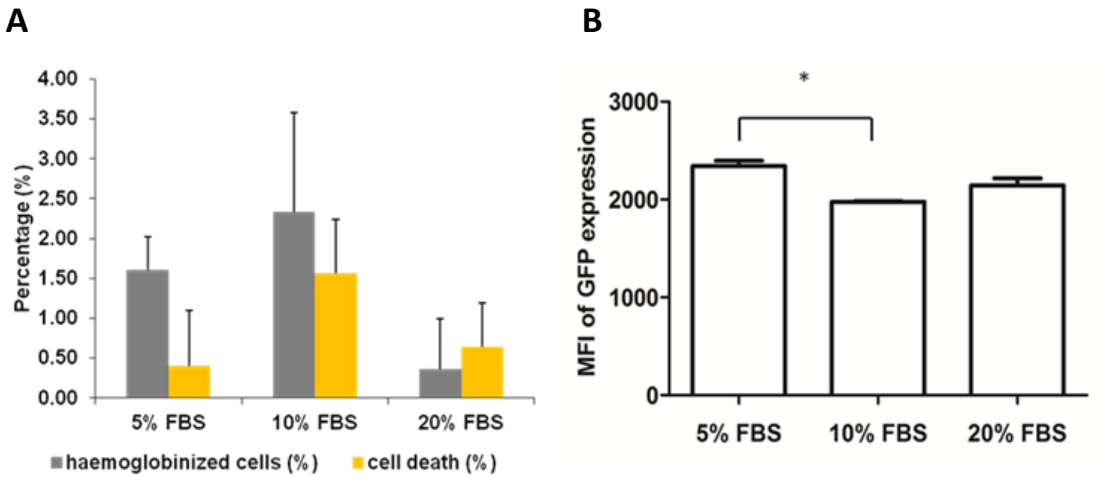


Figure 3.2.10. Analysis of spontaneous differentiation in βLCR-GFP-MEL cultures supplemented with different FBS concentrations. A) Values represent the percentage (%) of haemoglobinized cells (grey bars) and cell death (yellow bars) in cultures of βLCR-GFP-MEL cells supplemented with different concentrations of FBS, as indicated. Histograms represent mean ± SD for one experiment of triplicate sets. **B)** Values represent GFP mean fluorescence intensity (MFI-GFP) measured by flow cytometry on day 4 of culture. Histograms represent mean ± SD for one experiment of triplicates. Statistical analysis was performed by one-way ANOVA. Statistically different values (vs. control) are annotated (*, $p<0.05$).

Next, βLCR-GFP-MEL cells were cultured in the presence of 0.5%, 1%, 1.5%, and 2% DMSO for 4 days, followed by analysis for terminal erythroid differentiation using benzidine staining and for GFP expression by flow cytometry. Each condition was performed in triplicate. Previous work demonstrated that as soon as cells become committed to erythroid differentiation they will no longer require the presence of the inducer in order to complete their differentiation programme [252]. In addition, the proportion of cells that becomes committed in each cell generation was shown to increase proportionally to the concentration of inducer [252, 347]. Similarly, we showed that exposure of cells at high DMSO concentrations resulted in a large proportion of Hb-producing cells in the total population. As previously described [252, 347, 349], we reported that the erythroid-committed cells displayed gradual loss in size and condensed nuclei. We showed that induction with 1.5% and 2% DMSO resulted in 82% and 78% of haemoglobinized cells respectively, and thus concluded that either condition can be used for effective erythroid induction ($p>0.05$). Also, our results showed that induction with 0.5% and 1% DMSO resulted in significantly smaller proportions of haemoglobinized cells at 3% and 12.8% respectively.

Importantly, the proportion of haemoglobinized cells in cultures treated with 0.5% and 1% DMSO did not differ significantly from the proportion of spontaneously differentiated cells (~2%) in control cultures (**Figure 3.2.11A**). Following centrifugation of cells, we noted that the pellets of 1.5% and 2% DMSO-treated cultures were pink-to-red as a result of haemoglobin synthesis, whereas the pellets of control cultures were colourless (**Figure 3.2.11B**). These findings were in line with published data reporting that a 4-day induction with 2% DMSO resulted in erythroid differentiation at the normoblast stage, while extending cultures up to 7 days supported their differentiation into orthochromatophilic normoblasts. The authors reported synthesis of haeme at 48 h and of Hb at 72 h [344], indicating that a 4-day culture would suffice to activate *HBB* expression and accumulate HBB mRNA.

Moreover, we showed that as expected the growth rate of DMSO-treated cultures decreased with increasing concentration of inducer ($F(4,6)=82.31$, $p<0.0001$) (**Figure 3.2.11C**). Cultures of 0.5% DMSO showed similar growth rates to control cultures, exhibiting during end-point analysis 14–17-times the number of initially inoculated live cells. Cultures of 1% DMSO showed 10-fold multiplication in cell numbers while proliferation was 41% lower compared to control cultures. Cultures of 1.5% DMSO showed 3.5-fold multiplication in cell numbers while proliferation was 79% lower compared to control cultures. At 2% DMSO, cell numbers remained almost stationary, showing 1.8-fold multiplication in cell numbers while proliferation was 89.5% lower compared to control cultures (**Figure 3.2.11C**). Moreover, cultures treated with 2% DMSO exhibited the highest percentage of cell death at 7.2±0.4%. This was significantly higher than cell death for the other three conditions (0.0%–1.7%) ($F(4,7)=83.91$, $p<0.0001$), which showed similar cell death to control cultures (~1.6%) (**Figure 3.2.11A**).

Lastly, we monitored vector-derived *GFP* expression in cultures treated with DMSO by flow cytometry (**Figure 3.2.11D**). Induction with 1.5% and 2% DMSO produced comparable MFI-GFP (MFI; 8143.5 vs. 8311.86), which was ~4-times the basal *GFP* expression in uninduced control cultures (MFI; 1968.3), $p<0.0001$. Induction with 1% DMSO (MFI; 3997.5) produced approximately double the MFI-GFP compared to control cultures ($p<0.0001$), which however was approximately half that obtained with 1.5% or 2% DMSO treatment ($p<0.0001$). Induction with 0.5% DMSO (MFI; 2289.01) produced ~1.2-times the MFI-GFP of control cultures (**Figure**

3.2.11D). Importantly, we noted that the *GFP* expression profile for increasing concentrations of DMSO was in accordance with the percentage of haemoglobinized cells using benzidine staining.

An important experimental finding of this study was the observation that DMSO serves as a potent inducer of erythroid differentiation in GFP-MEL cells without detrimental effects on GFP expression. We thus chose DMSO at 1.5% instead of 2% concentration for subsequent experiments, because the former produced a higher percentage of erythroid-committed cells with minimal effect on viability and culture growth dynamics.

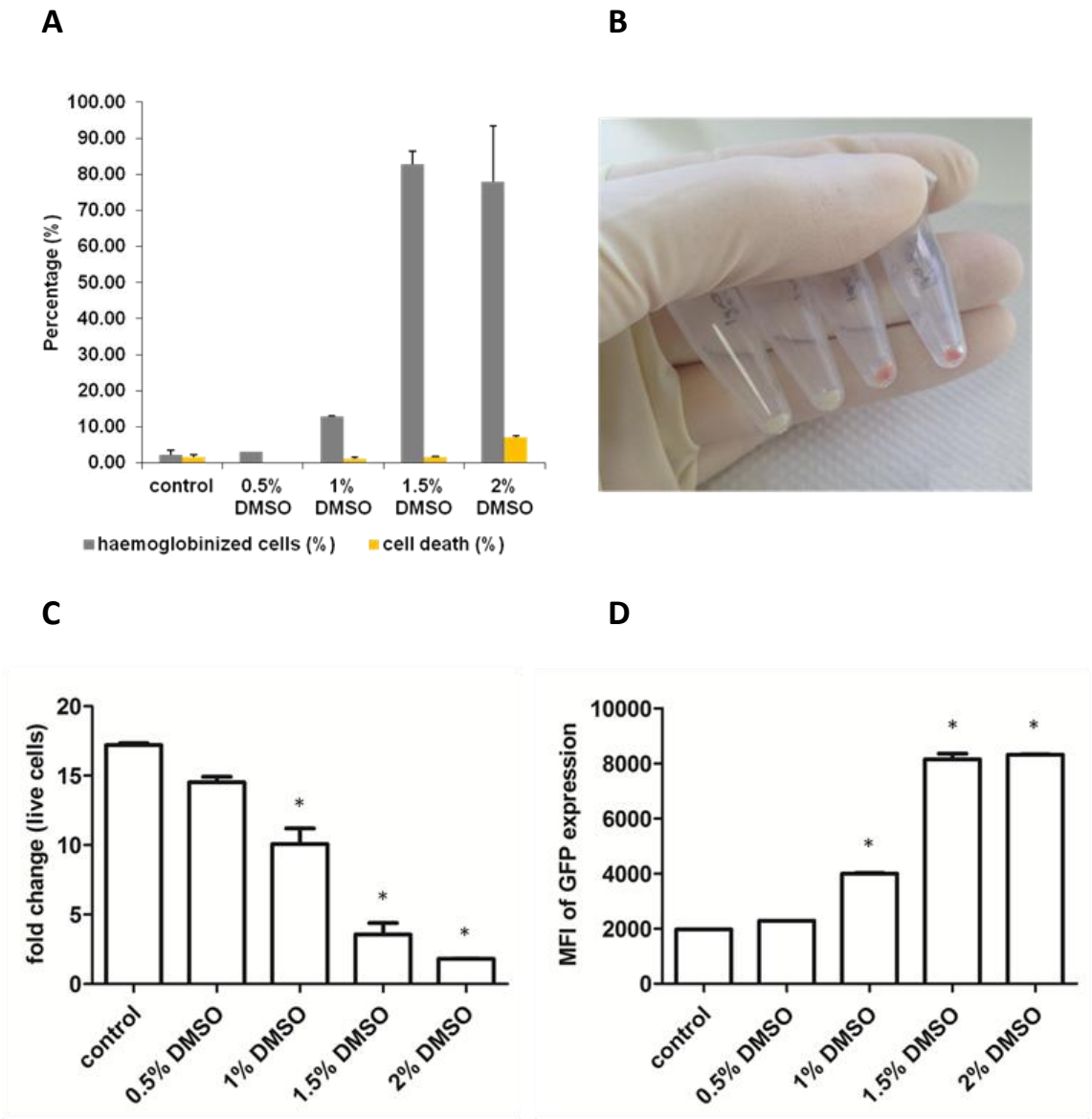


Figure 3.2.11. Analysis of erythroid differentiation in βLCR-GFP-MEL cultures induced with different DMSO concentrations. A) Values represent the percentage (%) of haemoglobinized cells (grey bars) and cell death (yellow bars) in cultures of βLCR-GFP-MEL cells supplemented with different concentrations of DMSO as indicated. The histogram represents mean ± SD. **B)** Pellets of βLCR-GFP-MEL cells treated with 2% DMSO (pink-to-red) or serving as controls (colourless). Image was captured as mobile photography. **C)** Values represent the fold change in the number of live cells at end-point cultures in comparison to the starting cell number. The histogram represents mean ± SD. **D)** Values represent GFP mean fluorescence intensity (MFI-GFP). The histogram represents mean ± SD. Analyses were performed on day 4 after induction. Each condition was performed in triplicate and statistical analysis was performed by one-way ANOVA after outlier removal. Statistically different values (vs. control) are annotated (*, p<0.05).

3.2.5.2 The effect of erythroid induction on GFP expression in GFP-MEL cells

We produced a PGK-GFP-MEL cell line demonstrating long-term GFP expression for the analysis of shRNA-triggered GFP knockdown. MEL cells were transduced and individual expanded clones were isolated by one round of limiting dilution to establish our GFP-MEL test cell line. Drawing on previous observations with GFP-HEL cells, we investigated whether induction of erythroid differentiation of GFP-MEL cells would affect GFP expression from the ubiquitous PGK promoter. The flow cytometric analysis of GFP-MEL cells separated the population into two distinct clusters on a FSC/SSC plot. We observed a shift of the SSC^{high}/FSC^{low} cells towards a SSC^{low}/FSC^{high} profile in cultures treated with 1.5% DMSO (**Figure 3.2.12A**). However, this observation was not statistically confirmed because the two sub-populations were not discretely separated on the density plot and therefore could not have accurately gated regions including the less frequent events of each sub-population. Nonetheless, the presence of the two sub-populations in similar patterns in naïve and induced cultures of MEL and GFP-MEL cells suggested that the cellular phenotype was not entirely dependent on erythroid induction per se but might have been influenced by additional variables. The MFI-GFP of each sub-population was measured by flow cytometry. To do so, regions were restricted to the central part of the sub-populations discernible on density blots, ignoring outliers and omitting areas of overlap between sub-populations. The resulting regions represented the vast majority of events for each sub-population of distinct FSC/SSC profile, facilitating powerful statistical analysis. The data shown represent four experiments of triplicate sets. Outlier removal was included in the analysis to reduce error variance and enhance accuracy of estimates. The correlation between erythroid induction and cell phenotype was highly significant, suggesting that both variables contributed to changes in MFI-GFP ($F(2,61)=8.07$, $p=0.0008$). Also, we detected a significant 18.7% reduction in MFI-GFP of the total cell population following treatment with DMSO vs. control cultures (MFI; 2683 vs. 3301, $p<0.001$). More specifically, we detected MFI-GFP loss by 76.3% following treatment with DMSO in the SSC^{high}/FSC^{low} sub-population vs. control cultures (MFI; 234 vs. 988, $p<0.001$), whereas MFI-GFP in the SSC^{low}/FSC^{high} sub-population did not change following erythroid induction vs. control cultures (MFI: 4814 vs. 4781) (**Figure 3.2.12B**). Under the assumption that cells are equally transduced,

monitoring of shRNA activity was performed by flow cytometry analysis for the total cell population.

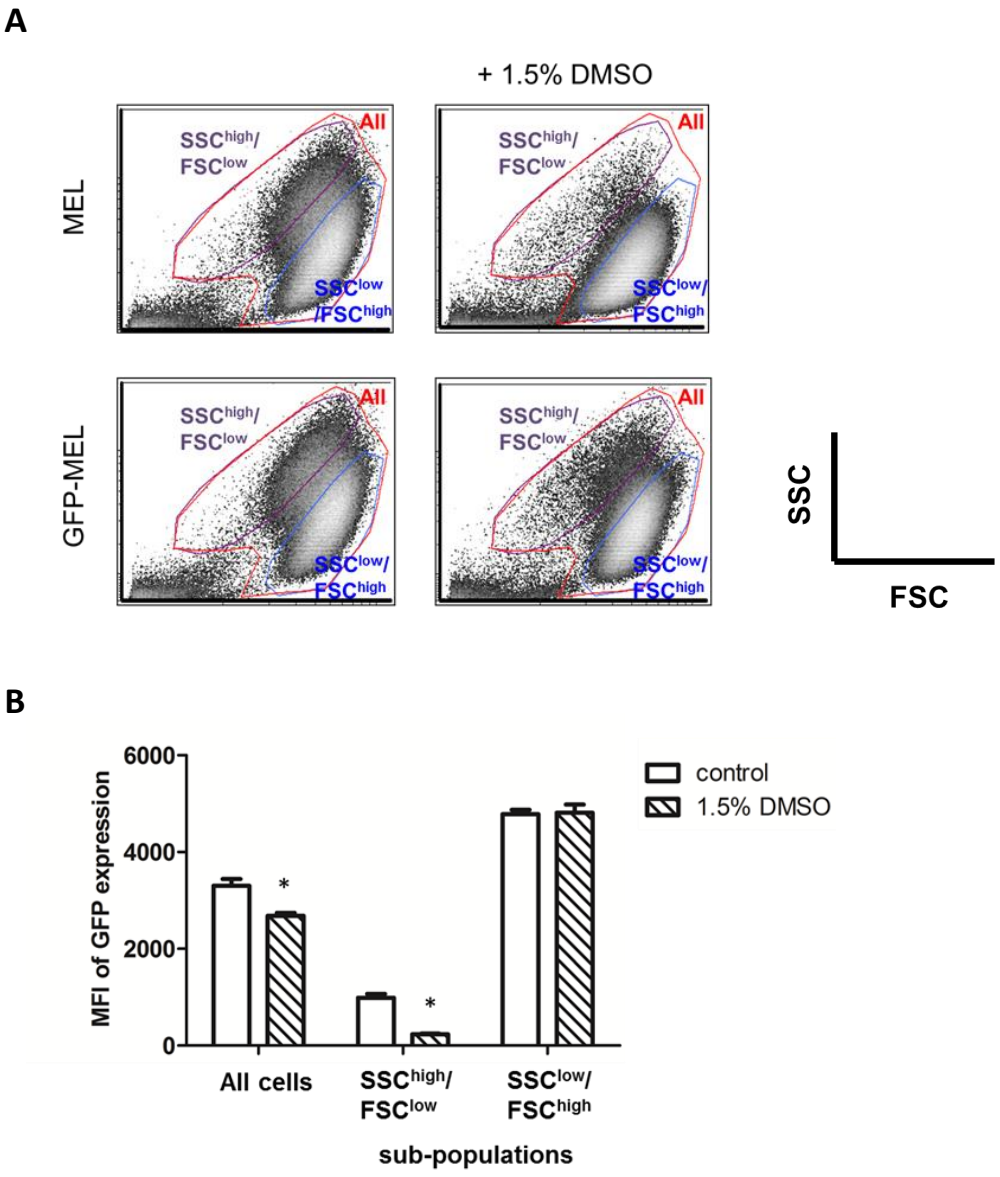
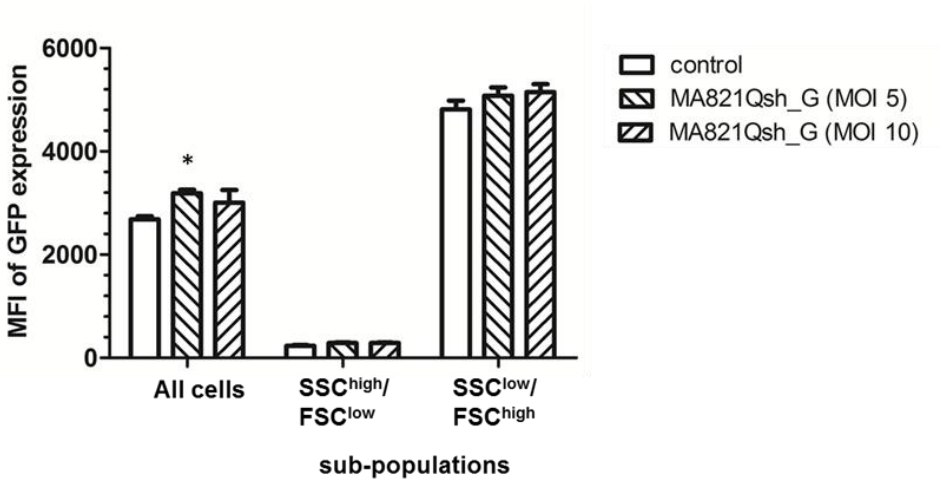


Figure 3.2.12. The effect of DMSO on PGK-GFP-MEL cell phenotype and GFP expression. A) Representative FSC vs. SSC dot plots of gated live cells in the total cell population (All cells) and each sub-population (SSC/FSC) separately for naïve MEL and GFP-MEL cells, in control and 1.5% DMSO-treated cultures. **B)** The histogram represents values of GFP mean fluorescence intensity (MFI-GFP) for the total cell population (All cells) and each sub-population (SSC/FSC) separately. Values are mean ± SD for four experiments of triplicates. Statistical analysis was performed by two-way ANOVA. Statistically different values (vs. control) are annotated (*, $p<0.05$).

3.2.5.3 Assessment of GFP-shRNA activity in erythroid-induced GFP-MEL cell cultures

In this study, PGK-GFP-MEL cells were transduced with the MA821Qsh_G LV and stabilised in cultures for 30 days to obtain a homogenous pool of clones. Considering that MEL cells are easily transduced by retroviruses [350] as well as the possibility of having a sub-optimal target knockdown, we performed parallel transductions using an MOI of 5 and 10, respectively. Transduced PGK-GFP-MEL cells were cultured in 1.5% DMSO-supplemented growth medium for 4 days, followed by analysis for GFP expression by flow cytometry. The data shown represents four experiments of triplicate sets after outlier removal. All cultures exhibited similar viability scores (>90%; data not shown) and >85% positivity for benzidine staining (data not shown). Importantly, we observed an absence of reduced MFI-GFP signal in cultures transduced with either MOI 5 or MOI 10, in the total cell population and the two sub-populations separately (**Figure 3.2.13**). Following analysis of GFP-MFI in the total cell population, in comparison to the control cultures (MFI; 2682.7±198.7), we detected a significant increase by 18.8% in the MFI-GFP of cultures transduced with MOI 5 (MFI; 3185.6±220.6) and a non-significant increase by 12.2% in the MFI-GFP of cultures transduced with MOI 10 (MFI; 3009.6±801.2) (**Figure 3.2.13**). These results corroborated previous findings with the GFP-HEL system, suggesting that the RNAi strategy was ineffective to produce stable and long-term knockdown in GFP reporter gene expression.

We next measured the expression of mRNA encoding HBB. Analyses were performed for two independent experiments of triplicate sets (**Figure 3.2.14A and 14B**). The primers were designed to specifically amplify HBB mRNA. We detected accumulation of total HBB transcripts in both MOI 5 and MOI 10 cultures, indicating proper RNA splicing of the lariat intron. In the 1st experiment (**Figure 3.2.14A**), we detected an average VCN of ~100 in bulk samples for both transductions by qPCR. The absence of detectable losses in MFI-GFP with multiple integrants per cell raised concerns over the functionality of the RNAi strategy. These findings were replicated in the 2nd experiment (**Figure 3.2.14B**).



			MFI-GFP	
			MFI	Δ% (fold change)
All cells	MA821Qsh_G	control	2682.7	
		MOI 5	3185.7	18.75 (1.19)
		MOI 10	3009.6	12.18 (1.12)
high SSC/ low FSC	MA821Qsh_G	control	234.1	
		MOI 5	285.8	22.08 (1.22)
		MOI 10	289.2	23.55 (1.24)
low SSC/ high FSC	MA821Qsh_G	control	4814.1	
		MOI 5	5078.7	5.49 (1.05)
		MOI 10	5147.8	6.93 (1.07)

Figure 3.2.13. Flow cytometry analysis of GFP expression in PGK-GFP-MEL cells transduced with the MA821Qsh_G LV at a MOI of 5 and 10. Cells were analysed for GFP mean fluorescence intensity (MFI-GFP) by flow cytometry on day 4 post-transduction. The histogram represents mean ± SD for four experiments of triplicate sets. The table shows MFI-GFP, the percentage change (Δ%) and fold change in MFI-GFP between the transduced and the control cultures for the total cell population (All cells) and each sub-population (SSC/FSC) separately. Statistical analysis was performed by two-way ANOVA after outlier removal. Statistically different values (vs. control) are annotated (*, p<0.05).

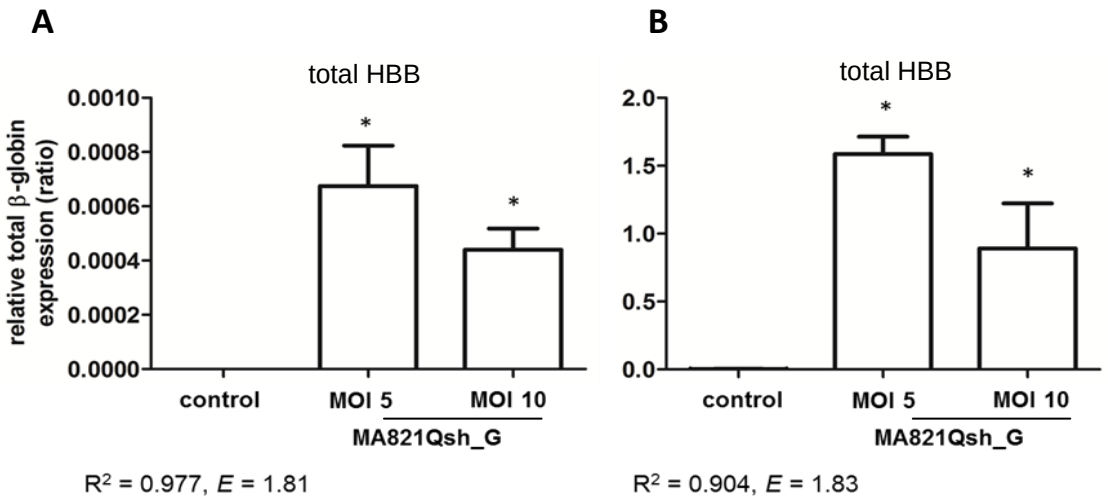


Figure 3.2.14. RT-qPCR analysis of mRNA expression in PGK-GFP-MEL cells transduced with the MA821Qsh_G LV at a MOI of 5 and 10. Cells were analysed for total HBB (endogenous and vector-derived) mRNA levels by RT-qPCR on day 4 post-transduction. Values are normalized by *GAPDH* expression. Statistical analysis was performed by one-way ANOVA. Statistically different values (vs. control) are annotated by (*, $p<0.05$). E: PCR amplification efficiency; R^2: square of Pearson's correlation coefficient.

3.2.5.4 Analysis of GFP knockdown using U6-driven shRNA in GFP-MEL cells

Up to this point, experimental data demonstrated proper splicing and transcription of the *HBB* transgene, which confirmed that insertion of the shRNA within the IVS-II region of the MA821Qsh_G vector did not compromise processing of the primary transcript. However, our results did not confirm whether the inserted shRNA was being efficiently excised and processed to trigger effective target knockdown. Based on these observations, we considered the possibility of having debilitated processing of the spliced lariat intron by the RNAi machinery as well as a non-optimal shRNA target sequence to guide gene-specific knockdown. Considering that the shRNA stem sequences had previously shown strong silencing activity [302], we next investigated the suitability of the loop sequence in our shRNA design, which differed from that published, for Dicer cleavage of the terminal loop and generation of potent duplex dsRNAs [209]. A total of three GFP-specific shRNAs were designed, all with identical stem but different loop structures (**section 3.1.5.1**). For the assessment of their functionality, direct assays exist, which are based on a stem-loop primer RT-qPCR protocol and which quantitatively measure the

accumulation of siRNA duplexes [351-354]. In spite of having reportedly great sensitivity of detection and specificity for even small siRNA quantities, the use of this assay is limited by the lack of easy-to-use design guidelines and the need of thorough optimisation to avoid false or inconsistent results, which at this juncture in our project rendered its use implausible. We therefore turned to an alternate, indirect approach of confirming shRNA functionality involving U6 promoter-based expression. Many studies have demonstrated the suitability of the human U6 small-nuclear-RNA promoter to drive expression of shRNAs **(section 3.1.5)**. Therefore shRNAs working under this system will possess functional stem and loop sequences. In turn, if these same shRNAs were found to fail to work under the MA821Q-based system, this would point towards a faulty shRNA processing by the RNAi machinery within this *HBB* intron insertion site-based LV design.

The evaluation of U6-shRNAs was performed with the PGK-GFP-MEL cell clone. Transduced cells were cultured with 1.5% DMSO for 7 days, and analysed on a daily basis for *GFP* expression and percentage of GFP^+ cells by flow cytometry. Measurements were recorded from day 2 post-transduction. Cells were collected during end-point analysis for VCN quantification by qPCR. The results shown in **Figures 3.2.15** and **3.2.16** are representative of three independent experiments of duplicate sets of cultures and include removal of values for samples having average VCN of 0 per cell. All cultures exhibited >75% viability scores (data not shown) and >85% positivity to benzidine staining (data not shown), demonstrating efficient induced erythroid differentiation. In comparison to control cultures, we showed that transductions with the U6-shRNA vectors resulted in significant losses in MFI-GFP ($F(8,258)=67.14$, $p<0.0001$). The three U6-shRNAs exhibited larger GFP knockdown effects with increasing length of time, such that by day 7 the pLKOsh_Ga, pLKOsh_Gb and pLKOsh_Gc vectors produced an MFI-GFP reduction of 85%, 40.8%, and 83.6%, respectively (**Table 3.2.1A**). The variability in the knockdown efficiency between these vectors could be attributed to differences in accessibility of the dsRNA knockdown triggers (siRNAs) to their target molecules and the vector copies per cell influencing the amount of siRNA duplexes available for silencing. In addition, differences in the shRNA design could affect Dicer processing into the production of siRNA duplexes with different seed regions, possibly of sub-

optimal activity. Consequently, such variables, together with residual protein persisting for some time after mRNA depletion, could contribute to potent siRNAs exerting their RNAi knockdown activity over a wider temporal range (e.g.: 12–120 hours post-transfection [355]). In addition, we detected a non-significant decrease in MFI-GFP by ~15% in cultures transduced with the parental MA821Q-based vectors (**Table 3.2.1A**). As these vectors do not encode RNAi elements that can guide GFP knockdown, we reasoned that this loss of MFI-GFP was a result of silencing of the PGK promoter driving expression of *GFP* after erythroid differentiation of the cells [356].

Also, we observed that vectors exhibiting significant losses in MFI-GFP similarly exhibited greater reductions in the percentage of GFP^+ cells $F(8,258)=50.61$, $p<0.0001$ (**Table 3.2.1B**). In comparison to control cultures, end-point analysis on day 7 showed that transductions with the pLKOsh_Ga, pLKOsh_Gb, and pLKOsh_Gc vectors had reduction in GFP^+ cells by 18.5%, 5.5%, and 15.1%, respectively (**Table 3.2.1B**).

A

	MFI-GFP (±SD)						
	day 2	day 3	day 4	day 5	day 6	day 7	*n*
negative control	4931.6±315.6	4214.8±6935	4114.9±840.3	4313.4±762.5	4330.5±516.5	3785.4±668.0	6
control	3998.1±203.6	3501.9±181.2	3463.1±139.5	3925.0±643.3	4306.7±1083.3	3567.7±1314.5	6
MA821	4006.4±276.3	3518.7±222.3	3375.7±258.3	3381.5±1114.6	3489.8±1454.5	3122.3±1672.0	6
MA821Q	4115.8±123.7	3583±112.6	3309.2±173.9	3369.7±759.3	3702.7±1144.0	3006.7±1202.7	6
MA821Qsh_G	3893.4±362.2	3545.0±240.0	3246.4±92.1	3509.6±910.0	3959.3±1258.0	3587.9±1514.9	6
pLKOsh_Scr	4146.2±232.6	3413.3±210.1	3353.4±75.3	3563.5±679.4	3921.1±1304.3	3451.7±1390.3	6
pLKOsh_Ga	2350.1±590.3	1568.4±323.7	1076.2±306.1	905.1±284.9	795.7±292.3	523.2±125.5	6
pLKOsh_Gb	3170.3±787.3	2415.0±837.7	2031±739.8	2025.5±786.2	2316.8±991.7	2111.9±1168.3	6
pLKOsh_Gc	2899.6±220.0	1887.9±822.0	1342.2±100.8	923.1±194.2	853.5±247.1	583.8±212.0	4

	Δ% (fold change)					
	day 2	day 3	day 4	day 5	day 6	day 7
control	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
MA821	0.21 (1.00)	0.48 (1.00)	-2.53 (0.97)	-13.85 (0.86)	-18.97 (0.81)	-12.48 (0.88)
MA821Q	2.95 (1.03)	2.31 (1.02)	-4.45 (0.96)	-14.15 (0.86)	-14.03 (0.86)	-15.72 (0.84)
MA821Qsh_G	-2.62 (0.97)	1.23 (1.01)	-6.26 (0.94)	-10.58 (0.89)	-8.07 (0.92)	0.57 (1.01)
pLKOsh_Scr	3.70 (1.04)	-2.53 (0.97)	-3.17 (0.97)	-9.21 (0.91)	-8.95 (0.91)	-3.25 (0.97)
pLKOsh_Ga	-41.22 (0.59)**	-55.21 (0.45)***	-68.92 (0.31)***	-76.94 (0.23)***	-81.53 (0.18)***	-85.33 (0.15)***
pLKOsh_Gb	-20.70 (0.79)	-31.04 (0.69)	-41.35 (0.59)**	-48.40 (0.52)***	-46.20 (0.54)***	-40.80 (0.59)**
pLKOsh_Gc	-27.47 (0.73)	-46.09 (0.54)**	-61.24 (0.39)***	-76.48 (0.24)***	-80.18 (0.20)***	-83.64 (0.16)***

B

	% GFP⁺ cells (±SD)						
	day 2	day 3	day 4	day 5	day 6	day 7	*n*
negative control	98.4±0.4	96.2±.5.4	97.9±2.4	98.0±1.8	98.0±1.0	96.6±2.1	6
control	99.1±0.4	98.7±0.2	99.2±0.8	98.6±0.7	97.8±1.0	95.2±1.8	6
MA821	98.0±1.8	96.7±1.6	98.5±1.7	97.2±1.7	96.0±2.7	94.1±0.7	6
MA821Q	98.5±1.2	98.1±1.2	98.8±0.6	97.4±1.0	95.5±2.1	93.3±2.3	6
MA821Qsh_G	98.7±1.2	98.5±0.9	97.3±3.6	97.7±1.3	96.5±1.8	94.9±1.2	6
pLKOsh_Scr	98.3±0.3	97.3±2.2	98.2±2.1	98.1±1.1	96.6±1.1	95.4±1.6	6
pLKOsh_Ga	98.8±0.5	95.5±2.0	90.9±7.2	85.5±9.1	79.4±8.6	77.6±8.8	6
pLKOsh_Gb	98.8±0.7	98.0±0.9	98.4±1.1	96.2±1.7	92.7±3.8	89.9±1.8	6
pLKOsh_Gc	98.4±1.2	95.9±0.9	95.5±4.6	89.0±2.4	79.7±2.8	80.8±2.2	4

	Δ% (fold change)					
	day 2	day 3	day 4	day 5	day 6	day 7
control	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
MA821	-1.06 (0.99)	-2.00 (0.98)	-0.73 (0.99)	-1.46 (0.99)	-1.85 (0.98)	-1.11 (0.99)
MA821Q	-0.55 (0.99)	-0.63 (0.99)	-0.43 (1.00)	-1.24 (0.99)	-2.37 (0.98)	-2.02 (0.98)
MA821Qsh_G	-0.41 (1.00)	-0.23 (1.00)	-1.93 (0.98)	-0.95 (0.99)	-1.31 (0.99)	-0.27 (1.00)
pLKOsh_Scr	-0.77 (0.99)	-1.39 (0.99)	-0.99 (0.99)	-0.56 (0.99)	-1.21 (0.99)	0.24 (1.00)
pLKOsh_Ga	-0.27 (1.00)	-3.27 (0.97)	-8.40 (0.92)***	-13.29 (0.87)***	-18.76 (0.81)***	-18.50 (0.81)***
pLKOsh_Gb	-0.25 (1.00)	-0.75 (0.99)	-0.79 (0.99)	-2.41 (0.98)	-5.24 (0.95)*	-5.50 (0.94)*
pLKOsh_Gc	-0.73 (0.99)	-2.83 (0.97)	-3.68 (0.96)	-9.72 (0.90)***	-18.54 (0.81)***	-15.08 (0.85)***

Table 3.2.1. Flow cytometry analysis of GFP expression and the percentage of GFP$^+$ cells in transduced PGK-GFP-MEL cells. PGK-GFP-MEL cells were transduced with various LVs as indicated, cultured in the presence of 1.5% DMSO, and analysed for **A)** GFP mean fluorescence intensity (MFI-GFP) and **B)** the percentage (%) of GFP$^+$ cells by flow cytometry on days 2–7 post-transduction. Values are mean ± SD for three experiments of duplicate sets. *n* represents the number of samples analysed per treatment. The percentage change (Δ%) and fold change in MFI-GFP between the transduced samples and the control, which is arbitrarily set at 1.00, are shown. Statistical analysis was performed by two-way ANOVA. Statistically different values (vs. control) are annotated (**, p<0.01), (***, p<0.001). Negative control: uninduced, untransduced; control: induced, untransduced.

Below, we report findings for each of the three experiments separately. Previous studies with MEL or HEL cells with a U6-shRNA-(GFP-SFFV) vector at a MOI of 1 resulted in >99% GFP^+ cells at 4-days post-transduction [357]. In the 1st experiment, we used a high MOI to produce significant expression of shRNAs and therefore detect phenotypic changes even in cases of siRNAs with moderate effects. In the 2nd and 3rd experiments, we calculated the volume of vector supernatant added to the cells using information from initial analyses in order to achieve an estimated average VCN per cell close to 1. Analyses for GFP expression and percentage of GFP^+ cells were performed by flow cytometry and presented in **Figures 3.2.15** (1st experiment), **Figure 3.2.19** (2nd experiment), **and Figure 3.2.20** (3rd experiment). For simplicity, we next elaborated on the findings of one representative experiment (1st experiment; **Figures 3.2.15–3.2.18**).

In the 1st experiment, GFP-MEL cells were transduced at a MOI of 50, followed by a 7-day induction. Analysis of *GFP* expression was performed by flow cytometry, which exhibited gradual loss in MFI-GFP with increasing length of time for all DMSO-treated cultures, while constant *GFP* expression was detected for the naïve (uninduced, untransduced) cultures. The extent of MFI-GFP loss between transduced cultures was shown to differ significantly, ($F(5,12)=19.30$, $p<0.0001$). In comparison to control (induced, untransduced) cultures (MFI: 2779.3±22.6), transduction with MA821Qsh_G (MFI: 2432.4±34.9; VCN: 28±6.7) produced a significant loss in MFI-GFP by 12.5% during end-point analysis ($p<0.05$) (**Figure 3.2.15A**). Unexpectedly, transductions with MA821 (MFI: 1757±104.3; VCN: 32.3±7.2) and MA821Q (MFI: 1876.1±389.8; VCN: 3.1±1.7) also produced significant reductions in MFI-GFP by 36.8% and 32.5% respectively (**Figure 3.2.15A**). Moreover, transductions with the U6-shRNA vectors produced significant losses in MFI-GFP, $p<0.0001$. In comparison to control cultures (MFI: 2779.3±22.6), the U6-based vectors pLKOsh_Ga (MFI: 590.9±40.5; VCN: 0.2±0.1), pLKOsh_Gb (MFI: 725.5±30.2; VCN: 0.3±0.2), and pLKOsh_Gc (MFI: 439.8±23; VCN: 3.3±1.9) produced an MFI-GFP knockdown of 78.8%, 73.9%, and 84.2% respectively at end-point analysis (**Figure 3.2.15A**).

Surprised to have detected *GFP* knockdown with vectors not encoding RNAi elements, we next compared the initial findings to data from subsequent experiments. In the 2nd experiment, in

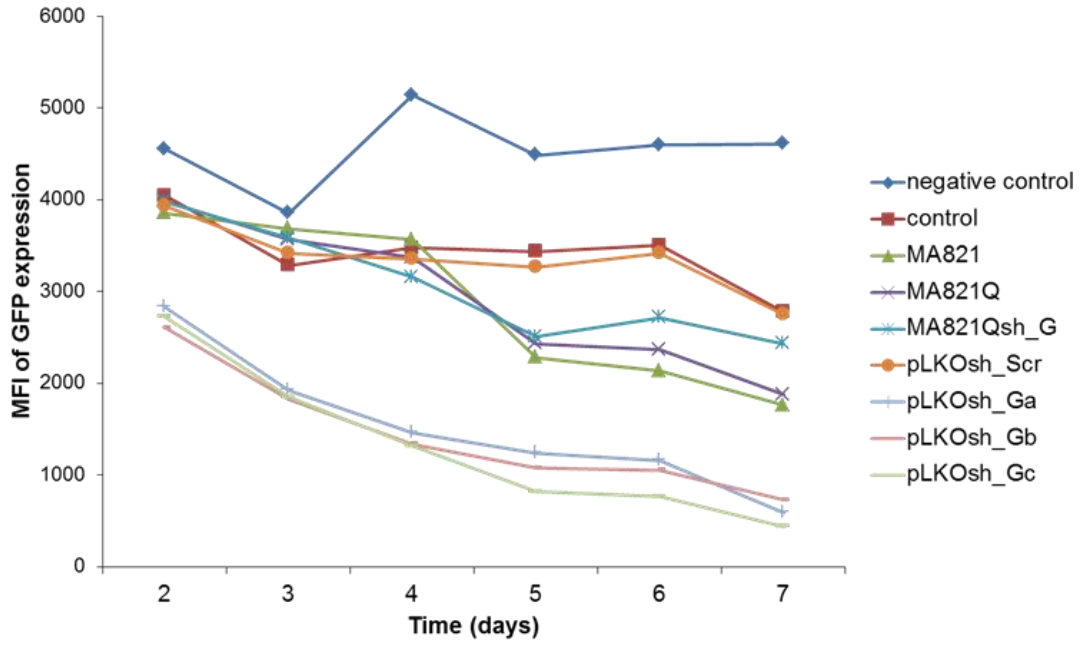
comparison to the control culture (MFI: 5256.9±255.3), transductions with MA821Qsh_G (MFI: 5529.2±145.9; VCN: 15.6±2.1) and MA821 (MFI: 5250.9±107.6; VCN: 22.7±6.2) did not affect MFI-GFP levels, whereas transductions with MA821Q (MFI: 4470.5±118.4; VCN: 44.0±6.4) produced significant reduction in MFI-GFP by 15% (**Figure 3.2.19A**). In the 3rd experiment, in comparison to the control culture (MFI: 2667±48.7), transductions with MA821Qsh_G (MFI: 2802.1±95.2; VCN: 4.9±3.4), MA821 (MFI: 2359±32.1; VCN = 96.6±7.6) and MA821Q (MFI: 2673.5±54.0; VCN: 53.7±0.3) did not affect MFI-GFP levels (**Figure 3.2.20A**). Overall, the initial findings of *GFP* knockdown for these vectors were not replicated in subsequent experiments in spite of using lower or higher average VCN per cell. Results suggested that the detectable change in MFI-GFP was in fact a result of stochastic fluctuations in gene expression occurring at random time intervals [358], as well as to inherent changes in *GFP* expression with the propagation of cells in culture. Alternatively, the scavenging of components of the transcription machinery by the strong *HBB* promoter could contribute towards weak *GFP* transcription, an effect which is not detected with the simpler U6 promoter of the pLKOsh_Scr vector. On the contrary, transductions with the U6-shRNA-GFP vectors produced significantly lower *GFP* expression in all three experiments (**Figures 3.2.15A, 19A and 20A**), indicating that the three shRNA structure designs are capable to trigger potent target knockdown.

In addition, we showed that vectors exhibiting a significant knockdown capability in MFI-GFP similarly exhibited greater reductions in the percentage of GFP$^+$ cells. In the case of the U6-shRNA based vectors, a decrease in MFI-GFP by ~79% correlated with a decrease in the percentage of GFP$^+$ cells by ~12% (**Figure 3.2.15B**). The reduction in MFI-GFP was also detected using fluorescence microscopy. Comparing the fluorescence microscope images of cultures transduced with MA821Qsh_G and pLKOsh_Ga vectors, which encoded the same shRNA, revealed a weaker MFI signal and a smaller number of fluorescent cells transduced with the latter (U6-based) vector (**Figure 3.2.16**). The MFI-GFP was quantified using an FL1 (GFP) histogram (**Figure 3.2.17**), based on histogram markers allowing for 5% false-positives for naïve MEL cultures as reference. At the initial stages of culture, we detected a clear peak of MFI-GFP, indicating uniform GFP expression in the total cell population. This profile of GFP expression persisted in cultures transduced with MA821Qsh_G, similarly to the control vectors.

However, over time we detected the presence of a second peak, shifted to the left of the first peak, in the panels of cultures transduced with pLKOsh_Ga and pLKOsh_Gb. Each peak represented a distinct cell cluster within the total cell population. On day 7, the two peaks exhibited similar height suggesting that 50% of the cells in the total population were expressing GFP at control MFI levels (~1500) and the other 50% of the cells were expressing GFP at a lower MFI (~800) (**Figure 3.2.17**). Cultures transduced with pLKOsh_Gc also displayed a second, lower peak, which over time became dominant in the population. From the plots, we estimated that ~80% of the cells in the population had been subjected to GFP knockdown to an MFI of ~800 (**Figures 3.2.17**). The proportion of cells in the transduced cultures exhibiting MFI-GFP at control levels could possibly not have been transduced, or transduction could have resulted in integrants within transcriptionally non-permissive genomic regions.

We then measured the expression of mRNA encoding GFP and HBB at the end of the 7 day period of induced erythroid differentiation (**Figure 3.2.18**). Despite a non-statistically significant outcome ($F(7,8)=2.33$, $p=0.1292$), transductions with the U6-based vectors pLKOsh_Ga (0.6±0.1; VCN: 0.2±0.1), pLKOsh_Gb (0.5±0.2; VCN: 0.3±0.2), and pLKOsh_Gc (0.4±0.1; VCN: 3.3±1.9) showed reduction in *GFP* expression by 47.4%, 50.1%, and 62.8% respectively (**Figure 3.2.15A and 18A**). Transduction with the MA821Qsh_G vector (1.1±0.2; VCN: 28±6.7) resulted in a similar expression of GFP mRNA to the control sample (1.1±0.3), whereas transduction with the non-shRNA containing MA821 vector (0.7±0.3; VCN: 32.3±7.2) led to a lower *GFP* expression by 32.4% (**Figure 3.2.15A and 18A**)**.** Lastly, in order to evaluate if the second intron of *HBB* containing the GFP shRNA cassettes within the MA821Qsh_G vector was being properly spliced, we measured relative *HBB* expression (**Figure 3.2.18B**). We detected expression of HBB mRNA in all cultures transduced with *HBB*-based vectors, which suggested proper splicing of intronic shRNA sequences.

A

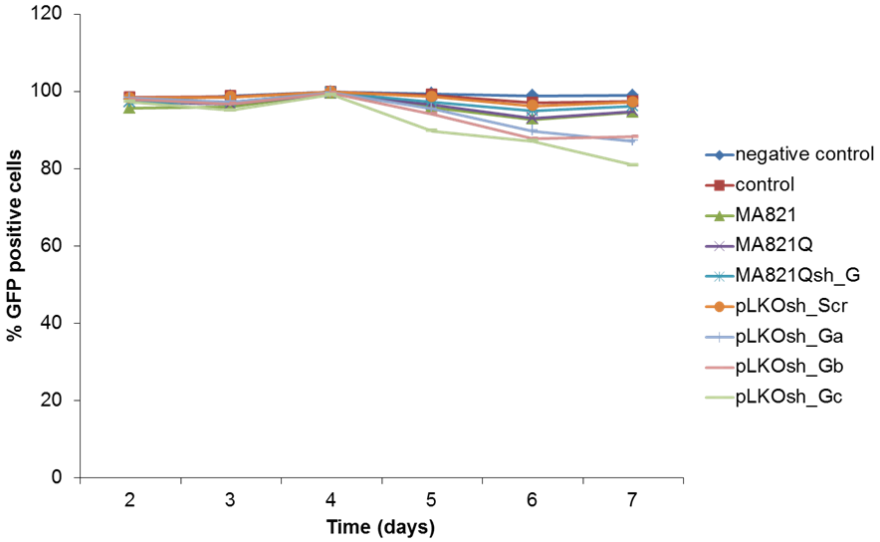


	MFI-GFP (±SD)					
	day 2	day 3	day 4	day 5	day 6	day 7
negative control	4553.3±19.2	3853.6±107.4	5139.4±51.4	4484.4±24.4	4595.5±49.4	4608.7±34.2
control	4046.2±167.4	3282.9±122.2	3474.7±113.9	3436.4±42.4	3500.9±68.8	2779.3±22.6
MA821	3852.1±13.9	3682.7±30.2	3564±58.1	2278.5±140.2	2132.7±133.0	1757±104.3
MA821Q	3987.8±117.9	3565±12.1	3368.2±185.7	2432.4±181.7	2362.3±408.4	1876.1±389.8
MA821Qsh_G	3976.2±195.4	3583.5±133.7	3155±50.0	2503.9±65.6	2716.1±33.5	2432.4±34.9
pLKOsh_Scr	3934.2±19.8	3416.8±17.3	3353.7±51.4	3264.9±9.3	3419.4±76.0	2752.4±36.0
pLKOsh_Ga	2830.3±11.9	1922±45.8	1460±14.5	1239.8±73.2	1159±68.0	590.9±40.5
pLKOsh_Gb	2605.2±67.9	1826.3±36.3	1338.9±57.3	1074.2±25.9	1051.8±23.2	725.5±30.2
pLKOsh_Gc	2725.1±8.5	1849.8±31.0	1313±6.5	817.6±24.3	997.8±4.4	439.8±23.0

	Δ% (fold change)					
	day 2	day 3	day 4	day 5	day 6	day 7
MA821	-4.80 (0.95)	12.18 (1.12)**	2.57 (1.03)	-33.69 (0.66)***	-39.08 (0.61)***	-36.78 (0.63)***
MA821Q	-1.44 (0.99)	8.59 (1.09)	-3.06 (0.97)	-29.22 (0.71)***	-32.52 (0.67)***	-32.50 (0.68)***
MA821Qsh_G	-1.73 (0.98)	9.15 (1.09)*	-9.20 (0.91)*	-27.13 (0.73)***	-22.42 (0.78)***	-12.48 (0.88)*
pLKOsh_Scr	-2.77 (0.97)	4.08 (1.04)	-3.48 (0.97)	-4.99 (0.95)	-2.33 (0.98)	-0.97 (0.99)
pLKOsh_Ga	-30.05 (0.70)***	-41.46 (0.59)***	-57.98 (0.42)***	-63.92 (0.36)***	-66.90 (0.33)***	-78.75 (0.21)***
pLKOsh_Gb	-35.61 (0.64)***	-44.37 (0.56)***	-61.47 (0.39)***	-68.74 (0.31)***	-69.96 (0.30)***	-73.90 (0.26)***
pLKOsh_Gc	-32.65 (0.67)***	-43.66 (0.56)***	-62.21 (0.38)***	-76.21 (0.24)***	-72.65 (0.27)***	-84.18 (0.16)***

	VCN (±SD)
negative control	0.0±0.0
control	0.0±0.0
MA821	32.3±7.2
MA821Q	3.1±1.7
MA821Qsh_G	28±6.7
pLKOsh_Scr	0.3±0.2
pLKOsh_Ga	0.2±0.1
pLKOsh_Gb	0.3±0.2
pLKOsh_Gc	3.3±1.9

B



	% GFP⁺ cells (±SD)					
	day 2	day 3	day 4	day 5	day 6	day 7
negative control	97.9±0.1	98.9±0.1	99.9±0.0	99.5±0.0	99±0.2	99±0.0
control	98.5±0.1	98.8±0.1	99.9±0.0	99.1±0.1	97.1±0.2	97.5±0.0
MA821	95.7±0.6	96.1±0.3	99.6±0.0	95.9±0.6	92.8±0.7	94.7±0.2
MA821Q	97.2±1	96.8±1.1	99.5±0.2	96.6±1.7	93.0±1.7	94.9±1.0
MA821Qsh_G	97.2±0.1	97.3±0.1	99.7±0.0	97.4±0.4	95.0±0.6	96.2±0.1
pLKOsh_Scr	98.3±0.3	98.6±0.2	99.9±0.0	98.7±0.1	96.3±0.1	97.3±0.0
pLKOsh_Ga	98.3±0.1	97.4±0.02	99.7±0.0	95.6±0.6	89.8±0.4	87.2±0.5
pLKOsh_Gb	98±0.1	96.8±0.1	99.6±0.0	94.2±0.4	87.8±0.2	88.3±0.4
pLKOsh_Gc	97.3±0.0	95.2±0.3	99.2±0.0	89.8±0.5	87.2±0.6	80.9±0.6

	Δ% (fold change)					
	day 2	day 3	day 4	day 5	day 6	day 7
MA821	-2.91 (0.97)***	-2.72 (0.97)***	-0.36 (1.00)	-3.21 (0.97)***	-4.37 (0.96)***	-2.89 (0.97)***
MA821Q	-1.33 (0.99)	-2.06 (0.98)***	-0.43 (1.00)	-2.52 (0.97)***	-4.16 (0.96)***	-2.66 (0.97)***
MA821Qsh_G	-1.38 (0.99)	-1.49 (0.99)*	-0.19 (1.00)	-1.76 (0.98)**	-2.12 (0.98)***	-1.26 (0.99)
pLKOsh_Scr	-0.24 (1.00)	-0.21 (1.00)	-0.05 (1.00)	-0.42 (1.00)	-0.83 (0.99)	-0.17 (1.00)
pLKOsh_Ga	-0.22 (1.00)	-1.41 (0.99)*	-0.20 (1.00)	-3.59 (0.96)***	-7.50 (0.92)***	-10.52 (0.89)***
pLKOsh_Gb	-0.58 (0.99)	-1.99 (0.98)**	-0.32 (1.00)	-4.97 (0.95)***	-9.52 (0.90)***	-9.41 (0.91)***
pLKOsh_Gc	-1.23 (0.99)	-3.63 (0.96)***	-0.70 (0.99)	-9.38 (0.91)***	-10.21 (0.90)***	-16.97 (0.83)***

Figure 3.2.15. VCN measurement, flow cytometry analysis of GFP expression and percentage of GFP$^+$ cells in transduced PGK-GFP-MEL cells – 1st experiment. PGK-GFP-MEL cells were transduced with various LVs as indicated, cultured in the presence of 1.5% DMSO and analysed for GFP mean fluorescence intensity (MFI-GFP) and the percentage (%) of GFP$^+$ cells by flow cytometry on days 2–7 post-transduction. **A)** The 2D-line graph shows MFI-GFP at different time points. Values are mean ± SD for one experiment of duplicate sets. The tables show MFI-GFP, the percentage change (Δ%) and fold change in MFI-GFP between the transduced samples and the control, as well as the average cell VCN in bulk samples. **B)** The 2D-line graph shows the % of GFP$^+$ cells at different time points. Values are mean ± SD for one experiment of duplicate sets. The tables show the percentage (%) of GFP$^+$ cells, and the percentage change (Δ%) and fold change in the proportion of GFP$^+$ cells between the transduced and the control cultures. Statistical analysis was performed by two-way ANOVA. Statistically different values (vs. control) are annotated (*, p<0.05), (**, p<0.01), (***, p<0.001). negative control: uninduced, untransduced; control: induced, untransduced.

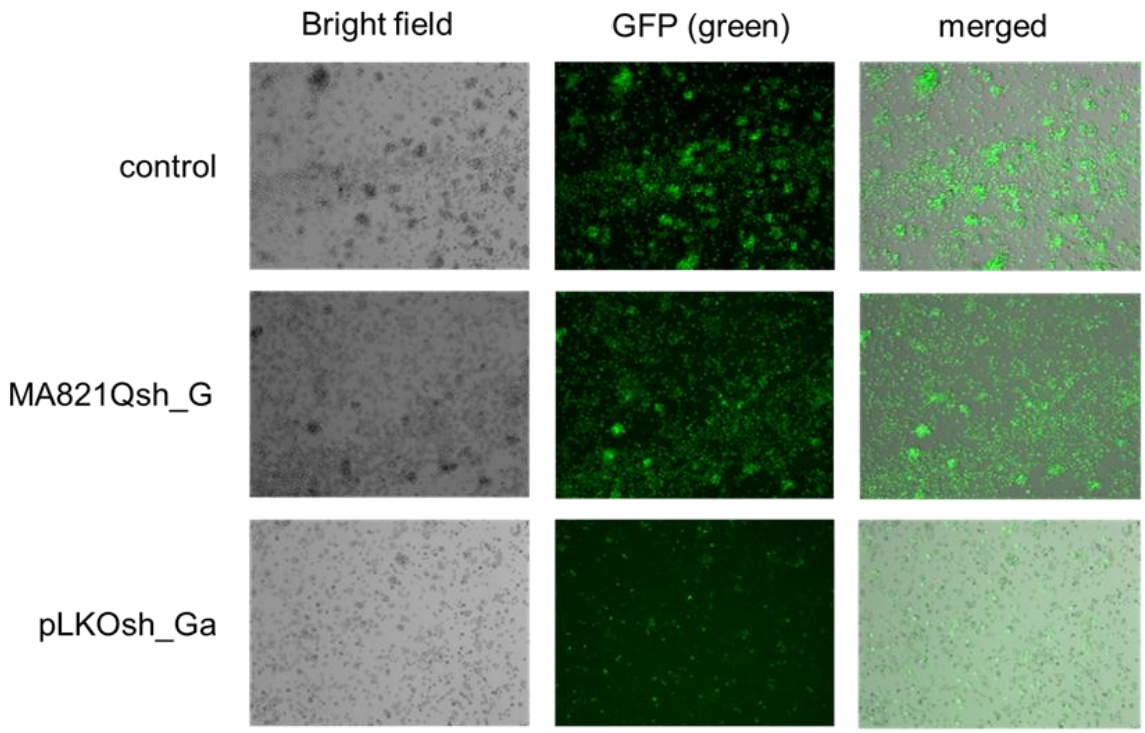


Figure 3.2.16. Assessment of GFP expression in transduced PGK-GFP-MEL cells by fluorescence microscopy. PGK-GFP-MEL cells were transduced with the MA821Qsh_G and pLKOsh_Ga LVs and induced to undergo erythroid differentiation with 1.5% DMSO. Both vectors encode identical shRNA structures. Images are captured by light microscopy (bright field) and fluorescence microscopy (GFP green) 7 days post-transduction. Merging of the two fields reveals the large number of GFP^+ cells exhibiting weak MFI in cultures transduced with pLKOsh_Ga. Magnification: 10x.

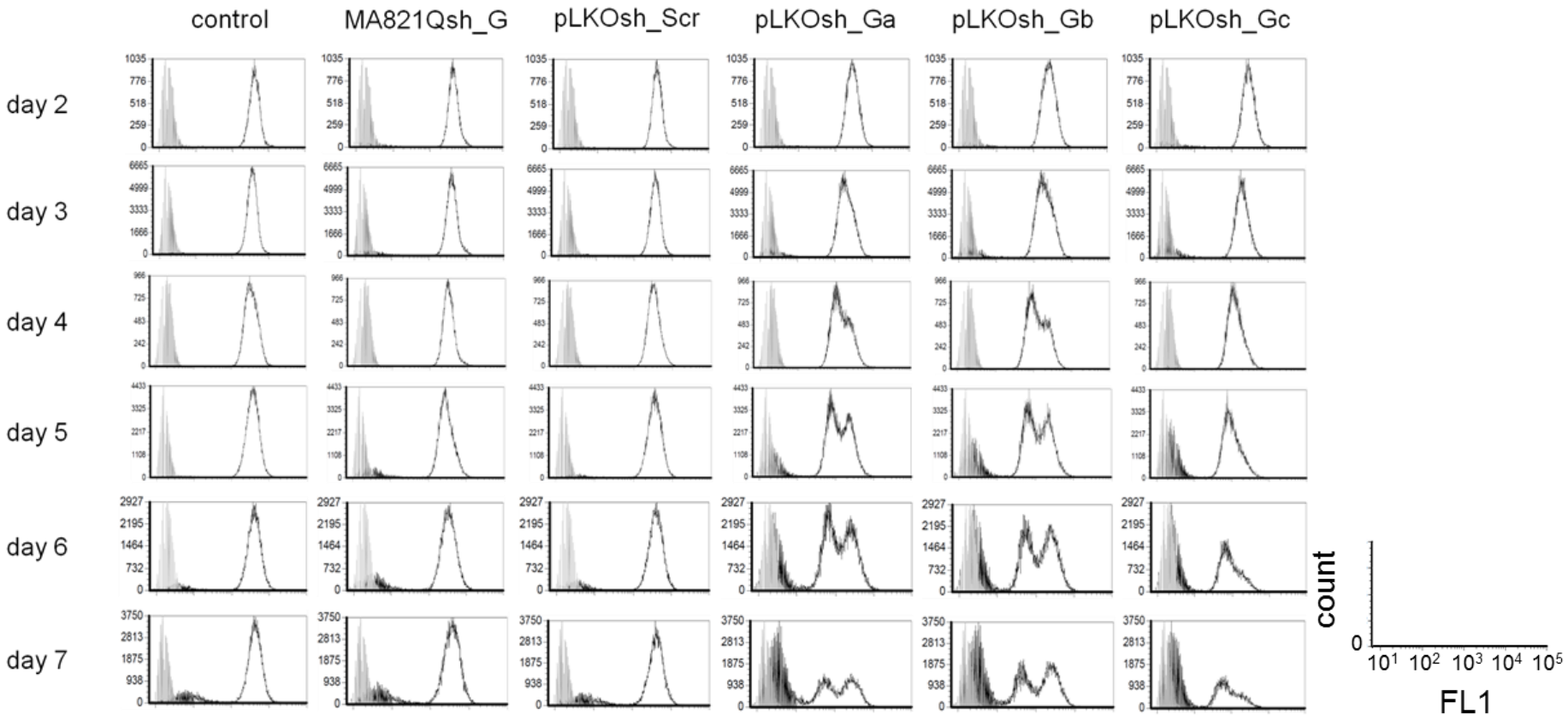
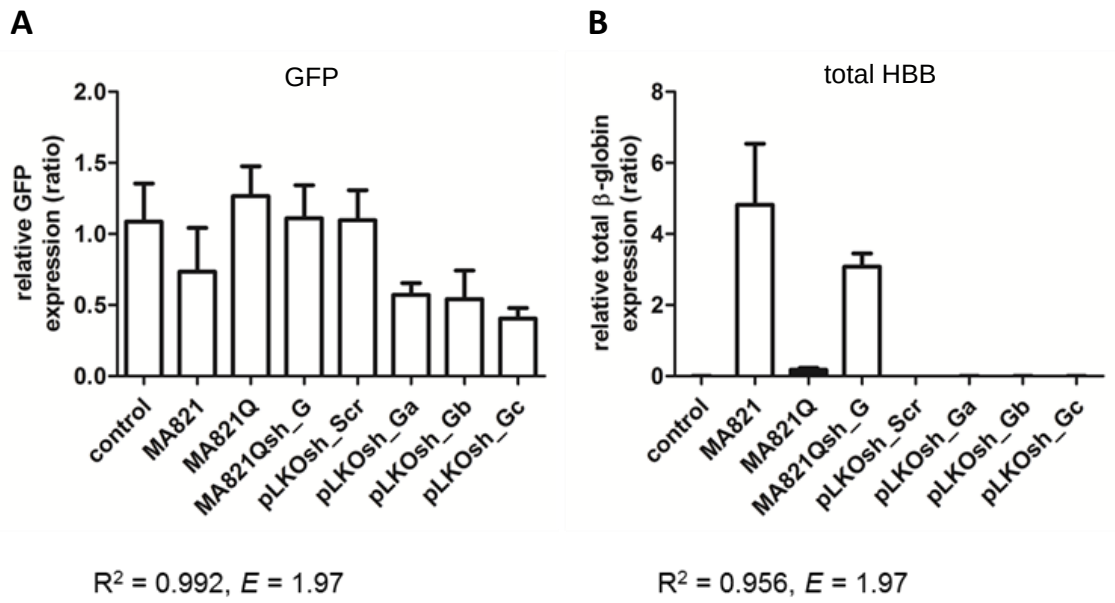


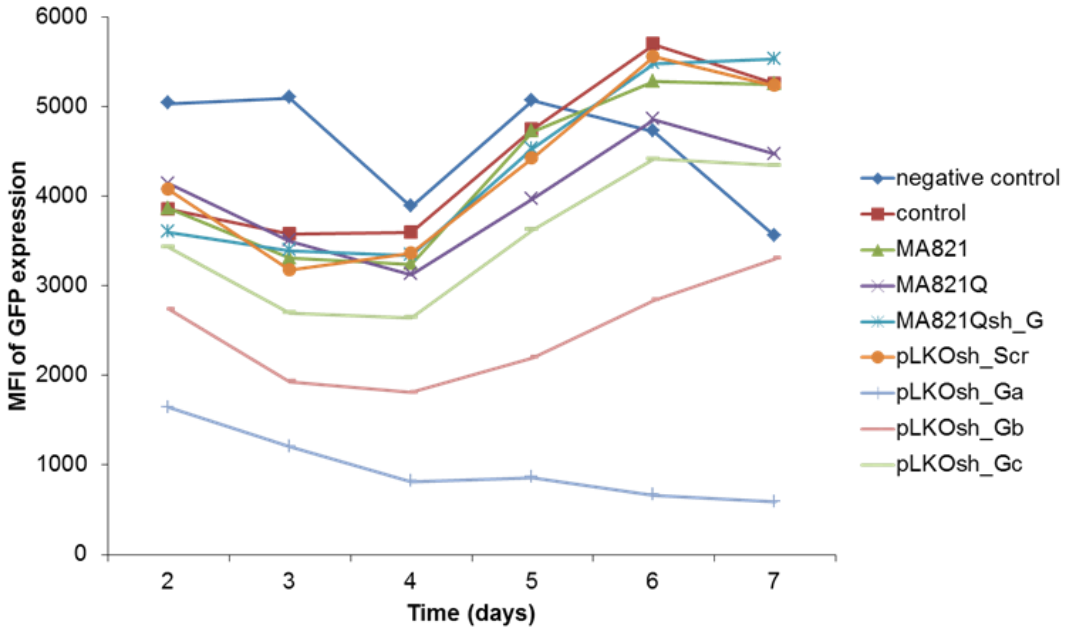
Figure 3.2.17. Flow cytometry analysis of GFP expression in PGK-GFP-MEL cells transduced with putative GFP knockdown vectors – count vs. FL1 plots. In each panel, naïve, GFP-MEL cells are shown as control (light grey outline) against test samples transduced with LV, as specified by labels in the top row (black outline).



	relative GFP expression	
	ratio (±SD)	Δ% (fold change)
control	1.1±0.3	
MA821	0.7±0.3	-32.39 (67.61)
MA821Q	1.3±0.2	16.64 (116.64)
MA821Qsh_G	1.1±0.2	2.23 (102.23)
pLKOsh_Scr	1.1±0.2	0.99 (100.99)
pLKOsh_Ga	0.6±0.1	-47.39 (52.61)
pLKOsh_Gb	0.5±0.2	-50.16 (49.84)
pLKOsh_Gc	0.4±0.1	-62.83 (37.17)

Figure 3.2.18. RT-qPCR analysis of mRNA expression in transduced PGK-GFP-MEL cells. PGK-GFP-MEL cells were transduced with various LVs as indicated, cultured in the presence of 1.5% DMSO, and analysed for **A)** GFP and **B)** total HBB (endogenous and vector-derived) mRNA levels by RT-qPCR on day 7 post-transduction. Values are normalized by *GAPDH* expression**.** Histograms represent mean ± SD for one experiment of duplicate sets. The table shows relative GFP expression, the percentage change (Δ%) and fold change in gene expression between the transduced and the control cultures. Statistical analysis was performed by one-way ANOVA. Statistically different values (vs. control) are annotated (*, p<0.05). E: PCR amplification efficiency; R^2: square of Pearson's correlation coefficient.

A

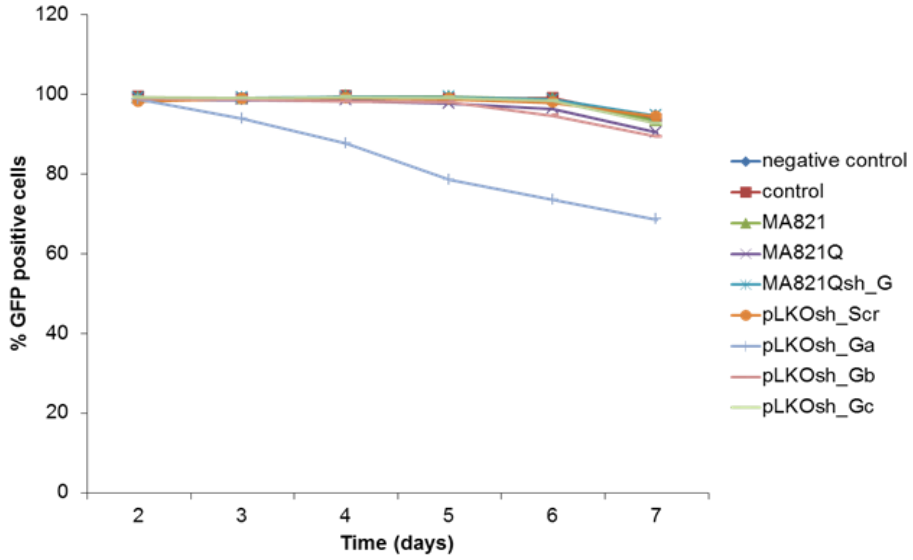


	MFI-GFP (±SD)					
	day 2	day 3	day 4	day 5	day 6	day 7
negative control	5037.2±195.4	5097.1±69.9	3883.8±211.6	5066.7±47.1	4722.6±82.3	3554±196.8
control	3854.2±277.6	3577.8±19.5	3596.1±17.0	4747.1±107.8	5696.4±70.6	5256.9±255.3
MA821	3860.8±313.8	3309.2±314.2	3242.2±465.9	4725.6±153.4	5282.9±71.9	5250.9±107.6
MA821Q	4140.6±78.8	3499.6±165.7	3126.6±111.1	3969.1±375.9	4859.2±5.4	4470.5±118.4
MA821Qsh_G	3598.8±399.2	3392.6±438.8	3338.5±70.2	4532.3±134.0	5483.1±103.7	5529.2±145.9
pLKOsh_Scr	4071.9±71.9	3177.0±6.9	3362.6±154.4	4426.7±33.2	5559.3±196.9	5231.2±1.2
pLKOsh_Ga	1641.6±414.6	1201.4±2.5	815.9±75.3	857.8±71.7	664.5±101.7	585.8±147.9
pLKOsh_Gb	2736.7±274.7	1930.4±194.1	1811.0±77.8	2197.6±26.1	2842.6±261.7	3306.3±230.0
pLKOsh_Gc	3430.8±66.4	2699.3±150.2	2645.4±123.0	3624.1±87.6	4416.9±132.0	4344.6±417.1

	Δ% (fold change)					
	day 2	day 3	day 4	day 5	day 6	day 7
MA821	0.17 (1.00)	-7.51 (0.92)	-9.84 (0.90)	-0.45 (1.00)	-7.26 (0.93)	-0.11 (1.00)
MA821Q	7.43 (1.07)	-2.18 (0.98)	-13.06 (0.87)	-16.39 (0.84)**	-14.70 (0.85)***	-14.96 (0.85)**
MA821Qsh_G	-6.63 (0.93)	-5.17 (0.95)	-7.17 (0.93)	-4.53 (0.95)	-3.75 (0.96)	5.18 (1.05)
pLKOsh_Scr	5.65 (1.06)	-11.20 (0.89)	-6.49 (0.94)	-6.75 (0.93)	-2.41 (0.98)	-0.48 (1.00)
pLKOsh_Ga	-57.41 (0.43)***	-66.42 (0.34)***	-77.31 (0.23)***	-81.93 (0.18)***	-88.33 (0.12)***	-88.86 (0.11)***
pLKOsh_Gb	-28.99 (0.71)***	-46.04 (0.54)***	-49.64 (0.50)***	-53.71 (0.46)***	-50.10 (0.50)***	-37.10 (0.63)***
pLKOsh_Gc	-10.98 (0.89)	-24.55 (0.75)***	-26.44 (0.74)***	-23.66 (0.76)***	-22.46 (0.78)***	-17.35 (0.83)***

	VCN (±SD)
negative control	0.0±0.0
control	0.0±0.0
MA821	22.7±6.2
MA821Q	44.0±6.4
MA821Qsh_G	15.6±2.1
pLKOsh_Scr	0.3±0.4
pLKOsh_Ga	1.8±1.3
pLKOsh_Gb	0.9±0.2
pLKOsh_Gc	0.0±0.0

B

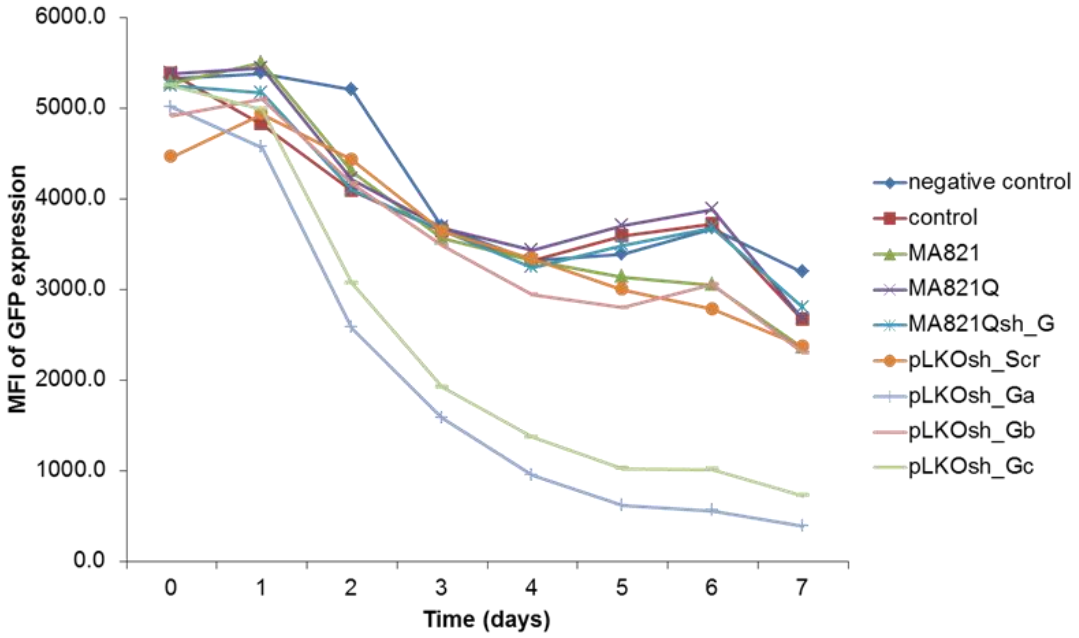


	% GFP⁺ cells (±SD)					
	day 2	day 3	day 4	day 5	day 6	day 7
negative control	98.7±0.1	98.6±0.0	99±0.1	98.9±0.1	98.1±0.2	94.5±0.5
control	99.3±0.0	98.8±0.1	99.4±0.0	98.8±0.9	99±0.2	93.6±0.3
MA821	99.2±0.1	98.8±0.1	99.5±0.0	99.8±0.1	98.8±0.1	93.4±0.8
MA821Q	98.8±0.0	98.6±0.1	98.4±0.1	97.6±0.1	96.3±0.4	90.4±0.7
MA821Qsh_G	99.2±0.2	99.1±0.0	99.3±0.1	99.2±0.1	98.7±0.2	94.7±1.1
pLKOsh_Scr	98.2±0.6	98.9±0.2	99.1±0.3	98.8±0.1	98±0.7	94.4±1.5
pLKOsh_Ga	98.7±0.2	93.9±2.6	87.7±3.7	78.5±9.7	73.5±6.3	68.6±5.9
pLKOsh_Gb	99±0.0	98.7±0.0	98.3±0.2	98.0±0.1	94.7±0.1	89.4±0.7
pLKOsh_Gc	99.2±0.0	99.1±0.1	99.3±0.0	99.1±0.1	98.4±0.1	92.6±0.2

	Δ% (fold change)					
	day 2	day 3	day 4	day 5	day 6	day 7
MA821	-0.08 (1.00)	-0.07 (1.00)	0.04 (1.00)	0.54 (1.01)	-0.21 (1.00)	-0.29 (1.00)
MA821Q	-0.48 (1.00)	-0.25 (1.00)	-1.02 (0.99)	-1.21 (0.99)	-2.69 (0.97)	-3.47 (0.97)
MA821Qsh_G	-0.09 (1.00)	0.31 (1.00)	-0.11 (1.00)	0.44 (1.00)	-0.27 (1.00)	1.12 (1.01)
pLKOsh_Scr	-1.05 (0.99)	0.04 (1.00)	-0.40 (1.00)	-0.09 (1.00)	-1.06 (0.99)	0.85 (1.01)
pLKOsh_Ga	-0.54 (0.99)	-4.99 (0.95)	-11.81 (0.88)***	-20.58 (0.79)***	-25.72 (0.74)***	-26.74 (0.73)***
pLKOsh_Gb	-0.27 (1.00)	-0.11 (1.00)	-1.16 (0.99)	-0.81 (0.99)	-4.37 (0.96)	-4.50 (0.96)
pLKOsh_Gc	-0.05 (1.00)	0.24 (1.00)	-0.15 (1.00)	0.29 (1.00)	-0.65 (0.99)	-1.06 (0.99)

Figure 3.2.19. VCN measurement, flow cytometry analysis of GFP expression and percentage of GFP$^+$ cells in transduced PGK-GFP-MEL cells – 2nd experiment. PGK-GFP-MEL cells were transduced with various LVs as indicated, cultured in the presence of 1.5% DMSO and analysed for GFP mean fluorescence intensity (MFI-GFP) and the percentage (%) of GFP$^+$ cells by flow cytometry on days 2–7 post-transduction. **A)** The 2D-line graph shows MFI-GFP at different time points. Values are mean ± SD for one experiment of duplicate sets. The tables show MFI-GFP, the percentage change (Δ%) and fold change in MFI-GFP between the transduced samples and the control, as well as the average cell VCN in bulk samples. **B)** The 2D-line graph shows the % of GFP$^+$ cells at different time points. Values are mean ± SD for one experiment of duplicate sets. The tables show the percentage of GFP$^+$ cells, and the percentage change (Δ%) and fold change in the proportion of GFP$^+$ cells between the transduced samples and the control. Statistical analysis was performed by two-way ANOVA. Statistically different values (vs. control) are annotated (*, p<0.05), (**, p<0.01), (***, p<0.001). negative control: uninduced, untransduced; control: induced, untransduced.

A

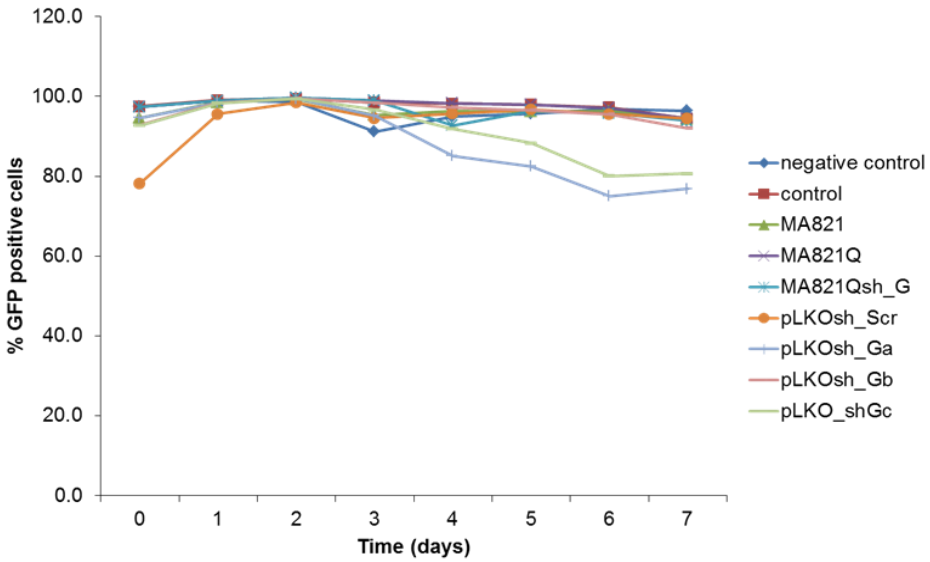


	MFI-GFP (±SD)							
	day 0	day 1	day 2	day 3	day 4	day 5	day 6	day 7
negative control	5335.0±96.1	5387.8±70.5	5204.3±46.9	3693.6±164.6	3321.7±135.0	3389.2±49.9	3673.6±116.0	3193.4±166.8
control	5382.5±83.5	48245±586.2	4093.7±194.4	3645.1±19.8	3318.5±81.2	3591.5±62.2	3722.9±124.0	2667±48.7
MA821	5280.6±155.1	5504±81.9	4306.1±115.9	3564.0±41.7	3320.8±24.5	3140.4±75.3	3053.9±246.0	2359±32.1
MA821Q	5383.9±56.1	5442.4±84.3	4219.1±32.8	3684.2±26.8	3432.6±15.4	3707.5±67.1	3886.6±203.7	2673.5±54.0
MA821Qsh_G	5248±149.1	5170.4±132.7	4105.3±425.6	3658.9±48.3	3245.7±37.3	3492.5±54.1	3678.7±93.1	2802.1±95.2
pLKOsh_Scr	4463.1±128.8	4939.8±179.2	4432.4±13.2	3646.2±10.3	3343.9±38.7	2998.9±20.9	2784.6±89.5	2371.4±137.3
pLKOsh_Ga	5012.4±125.6	4570.9±82.6	2578.4±32.6	1581.9±46.0	952.6±44.3	617.7±41.2	563.5±78.1	393.2±66.8
pLKOsh_Gb	4919.0±115.8	5101.4±14.8	4169.1±94.7	3488.4±51.5	2943.3±85.0	2804.7±78.5	3056.0±45.1	2304±1.9
pLKOsh_Gc	5252.9±4.9	4991.1±202.9	3074.1±153.0	1926.1±115.7	1371.5±164.4	1028.7±260.9	1020.3±268.2	727.8±226.7

	Δ% (fold change)							
	day 0	day 1	day 2	day 3	day 4	day 5	day 6	day 7
MA821	-1.89 (0.98)	14.07 (1.14)***	5.19 (1.05)	-2.22 (0.98)	0.07 (1.00)	-12.56 (0.87)*	-17.97 (0.82)***	-11.55 (0.88)
MA821Q	0.03 (1.00)	12.80 (1.13)***	3.06 (1.03)	1.07 (1.01)	3.44 (1.03)	3.23 (1.03)	4.40 (1.04)	0.25 (1.00)
MA821Qsh_G	-2.50 (0.98)	7.16 (1.07)	0.28 (1.00)	0.38 (1.00)	-2.20 (0.98)	-2.76 (0.97)	-1.19 (0.99)	5.07 (1.05)
pLKOsh_Scr	-17.08 (0.83)***	2.38 (1.02)	8.27 (1.08)	0.03 (1.00)	0.76 (1.01)	-16.50 (0.83)***	-25.20 (0.75)***	-11.08 (0.89)
pLKOsh_Ga	-6.87 (0.93)	-5.27 (0.95)	-37.02 (0.63)***	-56.60 (0.43)***	-71.29 (0.29)***	-82.80 (0.17)***	-84.86 (0.15)***	-85.26 (0.15)***
pLKOsh_Gb	-8.61 (0.91)*	5.73 (1.06)	1.84 (1.02)	-4.30 (0.96)	-11.31 (0.89)	-21.91 (0.78)***	-17.91 (0.82)***	-13.61 (0.86)
pLKOsh_Gc	-2.41 (0.98)	3.44 (1.03)	-24.91 (0.75)***	-47.16 (0.53)***	-58.67 (0.41)***	-71.36 (0.29)***	-72.59 (0.27)***	-72.71 (0.27)***

	VCN (±SD)
negative control	0.0±0.0
control	0.0±0.0
MA821	96.6±7.6
MA821Q	53.7±0.3
MA821Qsh_G	4.9±3.4
pLKOsh_Scr	13.8±9.3
pLKOsh_Ga	6.7±1.9
pLKOsh_Gb	2.0±0.2
pLKOsh_Gc	3.2±N/A

B



	% GFP⁺ cells (±SD)							
	day 0	day 1	day 2	day 3	day 4	day 5	day 6	day 7
negative control	97.5±0.0	98.9±0.3	98.6±0.1	91.2±8.2	94.9±0.1	95.7±0.1	96.9±0.5	96.4±0.1
control	97.5±0.2	99.1±0.1	99.4±0.2	98.5±0.2	98.2±0.4	97.9±0.2	97.3±0.3	94.4±0.2
MA821	94.7±0.1	98.5±0.0	99.2±0.1	95.3±0.3	96.3±0.2	96.2±0.1	96.3±0.2	94.3±0.0
MA821Q	97.4±0.1	98.9±0.0	99.6±0.0	98.9±0.2	98.3±0.3	97.9±0.1	97.0±0.6	94.5±0.2
MA821Qsh_G	97.3±0.2	98.9±0.2	99.7±0.0	98.9±0.1	92.7±0.9	96.4±0.3	95.8±0.1	93.8±0.5
pLKOsh_Scr	78.2±0.5	95.6±0.3	98.4±0.2	94.5±0.4	95.6±1.2	96.7±0.2	95.6±0.4	94.5±0.1
pLKOsh_Ga	94.6±0.2	98.5±0.1	99.4±0.0	95.1±0.7	85.1±0.5	82.5±0.8	75.0±1.9	76.9±0.4
pLKOsh_Gb	92.8±0.7	98.3±0.2	99.5±0.1	98.3±0.1	97.3±0.1	96.5±0.0	95.5±0.3	92.1±0.1
pLKOsh_Gc	92.7±2.5	98.3±0.2	99.4±0.2	96.6±0.8	91.8±3.0	88.3±3.8	80.1±4.8	91.9±0.3

	Δ% (fold change)							
	day 0	day 1	day 2	day 3	day 4	day 5	day 6	day 7
MA821	-2.95 (0.97)	-0.58 (0.99)	-0.21 (1.00)	-3.22 (0.97)	-1.88 (0.98)	-1.71 (0.98)	-1.02 (0.99)	-0.09 (1.00)
MA821Q	-0.13 (1.00)	-0.22 (1.00)	0.17 (1.00)	0.44 (1.00)	0.16 (1.00)	0.04 (1.00)	-0.26 (1.00)	0.10 (1.00)
MA821Qsh_G	-0.21 (1.00)	-0.20 (1.00)	0.23 (1.00)	0.49 (1.00)	-5.57 (0.94)**	-1.53 (0.98)	-1.57 (0.98)	-0.62 (0.99)
pLKOsh_Scr	-19.80 (0.80)***	-3.49 (0.97)	-1.01 (0.99)	-4.00 (0.96)	-2.56 (0.97)	-1.18 (0.99)	-1.75 (0.98)	0.06 (1.00)
pLKOsh_Ga	-3.01 (0.97)	-0.62 (0.99)	-0.05 (1.00)	-3.40 (0.97)	-13.29 (0.87)***	-15.76 (0.84)***	-22.90 (0.77)***	-18.57 (0.81)***
pLKOsh_Gb	-4.86 (0.95)*	-0.77 (0.99)	0.09 (1.00)	-0.13 (1.00)	-0.91 (0.99)	-1.42 (0.99)	-1.86 (0.98)	-2.47 (0.98)
pLKOsh_Gc	0.13 (1.00)**	-0.05 (1.00)	0.10 (1.00)	-2.91 (0.97)	-2.15 (0.98)***	-2.47 (0.98)***	-4.86 (0.95)***	-2.70 (0.97)***

Figure 3.2.20. VCN measurement, flow cytometry analysis of GFP expression and percentage of GFP$^+$ cells in transduced PGK-GFP-MEL cells – 3rd experiment. PGK-GFP-MEL cells were transduced with various LVs as indicated, cultured in the presence of 1.5% DMSO and analysed for GFP mean fluorescence intensity (MFI-GFP) and the percentage (%) of GFP$^+$ cells by flow cytometry on days 2–7 post-transduction. **A)** The 2D-line graph shows MFI-GFP at different time points. Values are mean ± SD for one experiment of duplicate sets. The tables show MFI-GFP, the percentage change (Δ%) and fold change in MFI-GFP between the transduced samples and the control, as well as the average cell VCN in bulk samples. **B)** The 2D-line graph shows the % of GFP$^+$ cells at different time points. Values are mean ± SD for one experiment of duplicate sets. The tables show the percentage of GFP$^+$ cells, and the percentage change (Δ%) and fold change in the proportion of GFP$^+$ cells between the transduced samples and the control. Statistical analysis was performed by two-way ANOVA. Statistically different values (vs. control) are annotated (*, p<0.05), (**, p<0.01), (***, p<0.001). negative control: uninduced, untransduced; control: induced, untransduced.

3.2.6 Discussion

RNAi is an evolutionarily conserved mechanism for gene silencing in which small dsRNA molecules target the destruction of cognate RNAs to elicit potent and selective target knockdown [197]. Specifically, stem-loop shRNA delivery exerts stable and long-lasting RNAi activity, and has been shown to work synergistically with globin gene delivery as a stem-cell-based therapy for sickle cell anaemia [224]. Here we attempted to extend the vector-encoded shRNA delivery system to develop a safe and highly efficient GT strategy for the treatment of β-thalassaemia. We describe the use of GFP-specific shRNAs to assess the efficiency of the shRNA insertion site as a means to quickly establish a proof of principle for our vector design.

In the present investigation, unexpected complexities and technical challenges were encountered on various efforts to establish a valid system for vector characterization, which prompted us to consider the use of three independent GFP-expressing mammalian cell-based systems, each in its turn. Initially, BM stem cells from an inbred GFP-transgenic mouse were harvested for vector characterization in haematopoietic CFC assays. After 14 days, clonal clusters of cells were classified *in situ* by light microscopy and recovered from the methylcellulose for analysis of GFP expression by flow cytometry. The success of this procedure was limited by the small number of cells per colony (≥30 cells [359]), which produced a fluorescence signal and a number of positive cells that fell below the detection threshold of the cytometer. The use of an *in vivo* spleen-colony assay represented an alternative attractive approach to generate a detectable fluorescence signal because the engraftment of HSCs in BM transplants would lead to production of macroscopic nodules in the spleen (CFU-S) consisting of up to a million cells and harvested for analysis [360, 361]. However, the need for optimization of BM transplantation procedures would have derailed our efforts to use GFP knockdown as a means of quickly establishing proof of principle for our shRNA expression strategy.

In consequence, and to ensure an abundance of material for analysis, clones of GFP-expressing HEL cells were produced. Optimal conditions for erythroid differentiation were established using haemin. These cells produced functional haemoglobin and expressed erythroid-specific markers in a similar fashion to differentiated primary human erythroid cells.

Detection of both HBG and HBB mRNA transcripts in cultures of differentiated GFP-HEL cells disagreed with published observations for an inactive *HBB* domain [250, 330, 334] and suggested that the HEL subline used in this study consisted of a transcription factor complement capable of providing adult *HBB* expression. However, it should be noted that the level of HBB polypeptide chains compared to those of HBG was not determined. Thus, although detectable, the absolute amount of this HBB protein may still be a small fraction compared to HBG in our HEL cell stock, which would still reflect a fetal globin expression pattern. In addition, the expression of erythroid-specific cell-surface markers in these cells did not correlate with the expression pattern expected over the course of erythroid lineage maturation. Therefore it would seem that HEL cells did not mirror the phenotype of primary erythroid cells, possible because many of their regulatory pathways have been modified over the process of transformation that led to their immortalisation [335]. Furthermore, the GFP-HEL cell line we employed as a model system was characterized by two distinct sub-populations and exhibited spontaneous phenotypic changes over long-term culture, which was further shifted towards a particular sub-population following treatment with haemin. Essentially, globin synthesis seemed to be unrelated to any phenotypic changes induced over the process of differentiation, such that our HEL subline could serve as an appropriate system for screening vectors with erythroid activity. Surprisingly, haemin induction of GFP-HEL cells profoundly reduced GFP expression (~98%), epistatic to any effects vector-induced silencing of GFP might have. Previous studies showing that the GFP fluorescence is not influenced by other proteins, co-factors or substrates [362], recommended the use of the GFP reporter for monitoring regulatory sequences. In our experiments, unexpected changes in GFP expression by induction of erythroid differentiation alone prompted us to consider the possibility of position effect variegation and heterochromatin-induced silencing of the integrated GFP-encoding cassette as a cellular response to haemin, requiring further experimentation in other GFP-expressing HEL cell clones. Alternatively, we considered the possibility that PGK promoter activity could be deregulated by haemin, requiring *de novo* construction of GFP-expressing HEL lines using vectors expressing GFP under alternate RNA-pol-II promoters. Although feasible, these experiments were beyond the immediate scope of our investigation. Assessment of vectors in undifferentiated GFP-HEL cells instead did not provide clear information on shRNA functionality, with non-detection of any

knockdown in expression, a result, however, that could be attributed to the reduced expression of erythroid-specific genes in the undifferentiated cells.

In response to these unexpected setbacks in the HEL system, clones of GFP-expressing MEL cells were produced. Optimal conditions for erythroid differentiation were established using DMSO. The effect of differentiation on PGK-driven GFP expression was associated with approximately 20% silencing, possibly due to the PGK promoter being prone to silencing over time [356]. In contrast to the GFP-HEL system, residual GFP expression in these cells seemed sufficient to measure the phenotype of moderate or weak siRNAs effectively. Resulting elevated expression of vector-derived HBB confirmed successful processing of its mRNA, while the failure to detect any effect of RNAi activity by quantification of RNAi target sequences in long-term cultures raised questions over the chosen shRNA expression strategy in our lentiviral vectors. The knockdown potency of the small dsRNA duplexes was confirmed using the U6 promoter-based system, also demonstrating that the loop structure consisting of 6 or 9 nucleotides in length was not influential on shRNA function. In conclusion from these findings it appeared that the failure to produce effective target knockdown was attributable to ineffective processing of the dsRNA hairpin from the spliced lariat intron by RNAi proteins.

In conclusion, experiments designed to target GFP as a speedy approach to assess shRNA vector functionality took longer than anticipated owing to difficulties in establishing a valid cell assay system. Taken together, our data in GFP-expressing cell-based systems demonstrated that the shRNA insertion site in the βIVS-II region did not affect post-transcriptional processing of the *HBB* transgene as shown by an increase in vector-derived *HBB* expression, which suggested proper splicing of intronic shRNA sequences. However, the failure to detect GFP knockdown from the MA821Qsh_G vector raised concerns over the functionality of the shRNA hairpin structure. Assessment of the shRNA design from a U6 promoter-based system confirmed the suitability of the stem and loop sequences for potent shRNA-mediated knockdown. Therefore, we proposed that the structure of the shRNA without additional regulatory sequences within the βIVS-II insertion site did not support correct processing of the hairpin by the RNAi proteins.

3.3 Functional evaluation of BCL11A-specific shRNAs in human cell lines

3.3.0 Aims

- To evaluate BCL11A-targeting LVs in suitable *in vitro* cell lines (K562 and HEL).
- To achieve HBG induction through BCL11A knockdown.

3.3.1 Introduction

Fetal haemoglobin (HbF) is a known modifier of the severity of HBB disorders. The ability of HbF to functionally compensate for deficient and/or abnormal HBB chain production was initially recognized in adult patients with coexisting mutations for benign hereditary persistence of fetal haemoglobin (HPFH), which exhibited milder clinical symptoms [59, 60]. The zinc-finger protein BCL11A is a key transcription factor involved in haematopoiesis that has recently been associated with HbF switching through genome-wide association studies [114, 115]. Follow-up investigations demonstrated that BCL11A functions as a repressor of *HBG1* and *HBG2* in adult erythroid cells and led to establishing BCL11A as a major transcriptional regulator of haemoglobin switching and silencer of HbF expression (**section 1.4.4.1**). On the basis of these findings, the treatment of β-thalassaemia through HbF re-activation via repression of BCL11A expression in adult erythroid precursors has presented a promising therapeutic approach.

This study, in response to the unexpected setbacks with the GFP-expressing HEL cell line (**section 3.2.3**) and in parallel to developments with the GFP-expressing MEL cell line (**section 3.2.5**), investigated the efficacy of LVs harbouring BCL11A shRNA-targeting moieties that through knockdown of this *HBG* repressor it was hoped to induce sufficient HBG chain production to be of therapeutic value in β-thalassaemia patients. Experiments were conducted with the human leukemia K562 and HEL cell lines, which display an embryonic and/or fetal pattern of HBB-like chain production.

3.3.1.1 Evidence for BCL11A expression in K562 cells

K562 is a human cell line derived from a plural effusion of a patient with chronic myelogeneous leukemia in blast crisis [251, 363]. A sub-clone of K562 cells was discovered to have an embryonic (*HBE*)/fetal (*HBG*) expression pattern and has been employed as a human erythroid model system. Although constitutively expressing *HBE* and *HBG*, K562 can be induced to differentiate into higher haemoglobin producing cells [251, 364, 365]. The induction of a K562-MEL hybrid cell line led to detectable levels of *HBB* expression, demonstrating that although K562 cells are trisomy for chromosome 11 and thus contain 3 copies of the *HBB* cluster [334] they do not possess the transcription factor complement necessary for *HBB* expression [364]. K562 cells have been widely used as a screening platform for identifying novel HbF inducers and studying the molecular mechanisms regulating *HBG* expression [366-368]. We therefore considered K562 cells as a suitable model to assess reactivation of *HBG* expression by targeting knockdown of its repressor BCL11A.

In human and mouse erythroid cells, BCL11A is expressed as several isoforms of which their individual role in globin gene expression is as yet unexplored. Different studies have presented Western blot analytical data indicating the presence of different BCL11A variants in K562 cells. A preliminary study [119] described a clone of K562 cells expressing only the short BCL11A-S isoforms (**Figure 3.3.1A**), similar to fetal liver erythroid cells that robustly express the *HBG* genes and in accordance to their embryonic-fetal globin expression phenotype. Another study [369] demonstrated moderate expression of median-sized BCL11A isoforms (**Figure 3.3.1B**). Others studies have found no expression of BCL11A in K562 cells [62, 370] whereas some reported the presence of the full-length and shorter isoforms [371]. It would seem that different clones of K562 cells contain a different BCL11A expression profile.

3.3.1.2 Evidence for BCL11A expression in HEL cells

HEL is a human erythroleukemia cell line developed from a patient who contracted leukemia after treatment for Hodgkin's disease [250]. [250]. HEL cells can be induced to differentiate into megakaryocytes, macrophages, or erythroid type cells [250, 372, 373]. Their differentiation down the erythroid pathway is characterized by globin polypeptide synthesis, mainly of the fetal HBG type with trace amounts of HBE chains but no HBB chains. HEL cells were shown to contain an intact *HBB* cluster in developmental arrest, either owing to the absence of activating or the presence of inhibiting *trans*-acting elements regulating *HBB* expression [330, 333, 374].

Published evidence for BCL11A expression in HEL cells is very limited. However, the anti-BCL11A antibody (*Santa* Cruz) used in our study for Western blot analysis, was functionally characterized in protein lysates from HEL cells and described in the manufacturer's brochure to bind the large protein variants.

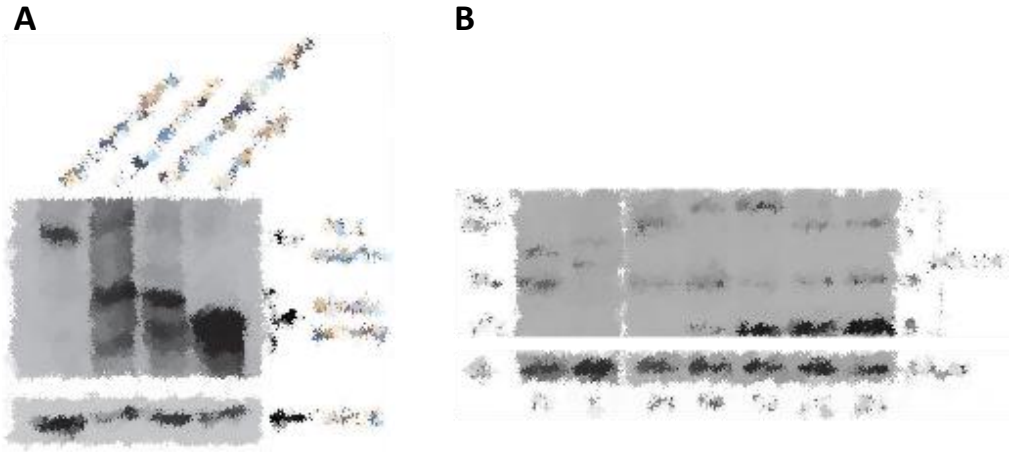


Figure 3.3.1. Detection of BCL11A isoforms in K562 cells by Western blot analysis. A) Cell lysates of human bone marrow (BM), fetal liver (FL), primitive erythroblasts and K562 were analysed by Western blotting and tested for the presence of BCL11A. GAPDH served as a loading control. Reproduced from [119]. **B)** Cell lysates of human fetal liver (FL), K562 (K) and adult $CD34^+$ cells cultured for 3–14 days (D3–D14), analysed by Western blotting and tested for the presence of BCL11A. β-Actin served as a loading control. Reproduced from [369].

3.3.2 Evaluation of BCL11A-targetting LVs in K562 cells

Before proceeding with large-scale experiments on all vectors, the transducibility of K562 cells with LVs was tested at increasing MOI (MOI 10, 25 and 50) with the aim to determine optimal conditions for detecting knockdown of *BCL11A* expression. Transductions were performed in singlet using one of the U6-based BCL11A-targeting LVs (BCL11A_449-shRNA) and the control vectors for RNA–pol-II (MA821Qsh_G) and RNA-pol-III (pLKOsh_Scr) shRNA-mediated expression. K562 cells were transduced (**section 2.8.4**) and grown for 7 days in the presence of 150 μM hydroxyurea to induce higher levels of HbF expression. [Note: our laboratory had previously established conditions for HbF induction in K562 cells using hydroxyurea, sodium butyrate and resveratrol. In this line of experiments, erythroid induction was performed with hydroxyurea since, in contrast to the other compounds [371, 375, 376], its molecular mechanism acted independently of the BCL11A regulatory pathway [377].] The fraction of erythroid differentiated cells was determined by benzidine. Quantification of BCL11A knockdown and HBG synthesis was performed by Western blotting.

The fraction of benzidine-positive cells in the control culture (induced, untransduced) was ~1.5 times higher compared to the undifferentiated culture (33.3% vs. 22.5%), indicating a low degree of erythroid differentiation upon induction and showing that K562 cells had spontaneously differentiated in serum-supplemented media as previously suggested [378]. The experimental procedure was non-toxic to cells exhibiting a low 5.99±1.90% of cell death. LV transduced cell cultures exhibited similar or lower fractions of benzidine-positive cells than the control (**Figure 3.3.2**). However, we showed that transduction with the U6 promoter-based pLKOsh_B9_(7-1a) vector led to higher number of haemoglobinized cells across different MOIs, with a maximum of 48% at MOI 50. The pLKOsh_B9_(7-1a) vector encodes a publically validated shRNA targeting BCL11A and could thus confirm the functionality of the shRNA design. Consequently, effective shRNA-mediated BCL11A knockdown would in principle lead to concomitant derepression of *HBG* expression and HbF production, resulting in a higher fraction of benzidine-positive cells.

Expression of BCL11A protein was assessed by Western blotting with lysates extracted from either $2x10^5$ or $4x10^5$ cells (**Figure 3.3.3**). These immunoblots showed a persistent background of bands, which were identified as the median-size (~51 kDa) and shorter (~35 kDa) isoforms of BCL11A, whereas, the larger isoforms (XL: 120 kDa, L: 100 kDa) were absent. Staining of the blots resulted in persistent background bands and weak BCL11A-specific signals that were difficult to interpret and quantify without artificial enhancement of their intensity. Consequently, we inferred only that transductions with pLKOsh_B9_(7-1a) gave similar BCL11A expression profiles to the control vectors.

As a result, we proceeded by quantifying expression of HBG protein in order to evaluate for BCL11A knockdown. The expression of HBG protein increased in the K562 cells transduced with pLKOsh_B9_(7-1a) and this result was conspicuous at larger vector doses (MOI 25 and 50). No increase in HBG protein was detected in cultures transduced with the HBB cassette-based MA821Qsh_B9 vector. On the contrary, these cultures showed reduction in HBG levels with a concomitant increase in the expression level of HBB polypeptides. This result was also found with the MA821Qsh_G vector, which serves as a control for RNA-pol-II shRNA-mediated expression. In comparison to the control non-transduced cultures, the enhanced HBB protein signal in cultures transduced with MA821Q-derived vectors was attributed to expression of the *HBB* transgene contained within this LV series of constructs. In subsequent experiments we interrogated the possibility of competition between the vector-derived *HBB* and the endogenous *HBG* for access to the transcriptional machinery.

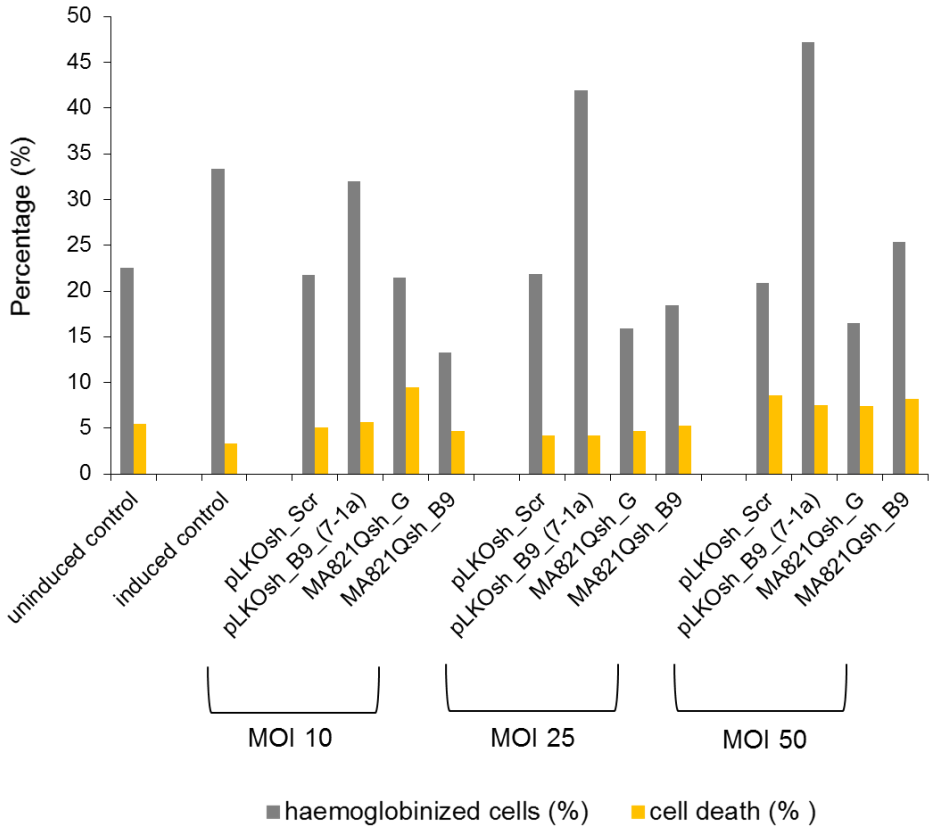


Figure 3.3.2. Erythroid induction of K562 cells. K562 cells were transduced with various LVs as indicated at a MOI of 10, 25 and 50, cultured for 7 days in the presence of 150 µM hydroxyurea, and analysed for the percentage (%) of haemoglobinization (grey bars) and cell death (yellow bars) by benzidine and trypan blue staining respectively. Data are presented for single transductions from one experiment.

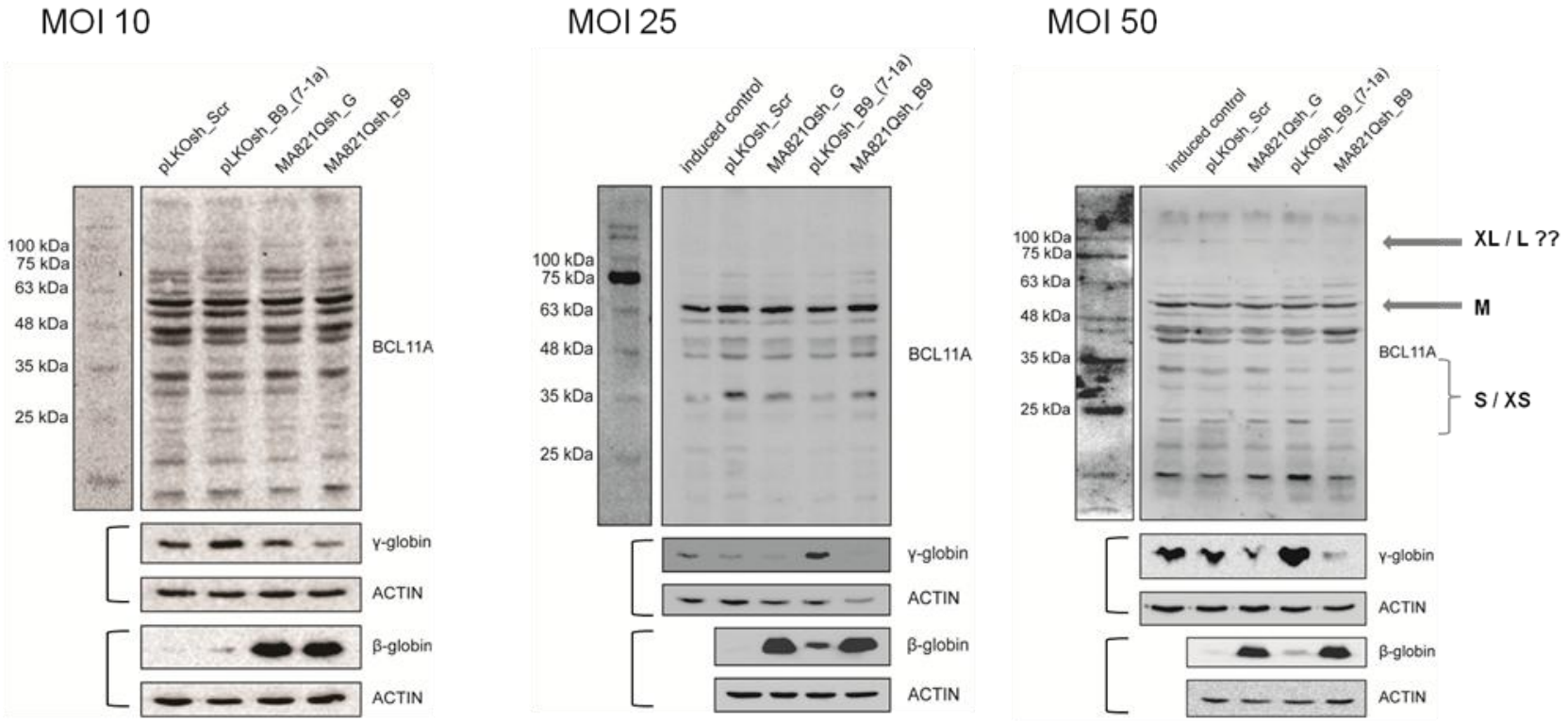


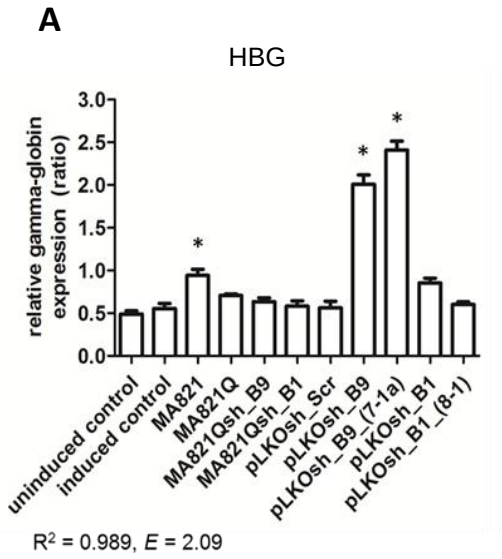
Figure 3.3.3. Expression levels of BCL11A, HBG and HBB protein in K562 cells by Western blot analysis. K562 cells were transduced with LVs as indicated at a MOI of 10, 25 and 50, blots were probed with antibodies against BCL11A (*Santa Cruz*), HBG, and HBB in whole cell extracts from $2x10^5$ cells (at MOI 10) and $4x10^5$ cells (at MOI 25 and 50). ACTIN was used as internal control. Molecular weight of BCL11A isoforms XL/L/M/S/XS: 125/100/51/35/25 kDa.

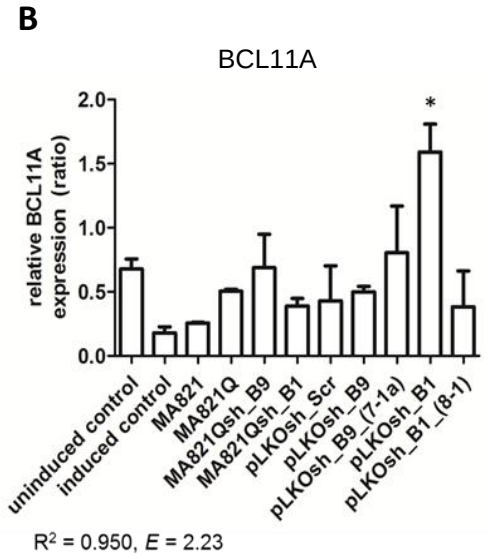
A large-scale experiment with all vectors was set-up with transductions performed in duplicate at a high vector dose (MOI 50), which had previously exhibited readily measurable effects on *HBG* and *HBB* expression. As before, we showed a relatively low degree of erythroid differentiation into haemoglobinized cells (15%–20%) with the exception of transductions with the U6 promoter-based pLKOsh_B9_(7-1a) and pLKOsh_B9 vectors, which led to a larger fraction of benzidine-positive cells of 51% and 36% respectively. Also, cultures transduced with pLKOsh_B9_(7-1a) and pLKOsh_B9 led to an increased expression of HBG mRNA by 334.7% (2.41±0.2) and 262.7% (2.01±0.2) (p<0.0001) respectively, compared to the control culture (0.55±1.1) (**Figure 3.3.4A**). There was no increase in the level of *HBG* expression observed in transductions with the U6-based pLKO.1 BCL11A_451-shRNA vectors (**Figure 3.3.4A**). RT-qPCR analysis of BCL11A expression was inconclusive since primers amplifying different BCL11A siRNA target sequences gave different results; one primer set (shB9) failed to produce detectable amplicons suggesting absence of the targeted BCL11A region in the K562 transcriptome, while the other (shB1; **Figure 3.3.4B**) gave BCL11A mRNA levels comparable to those of control cultures. Both primer sets were designed to amplify BCL11A regions, which were common to all known BCL11A transcript variants annotated by *Ensembl* (on-line at http://www.ensembl.org/index.html) (**Figure 3.3.5**)**.** The specificity of the primers for their target sequence was confirmed in samples of human erythroid progenitors.

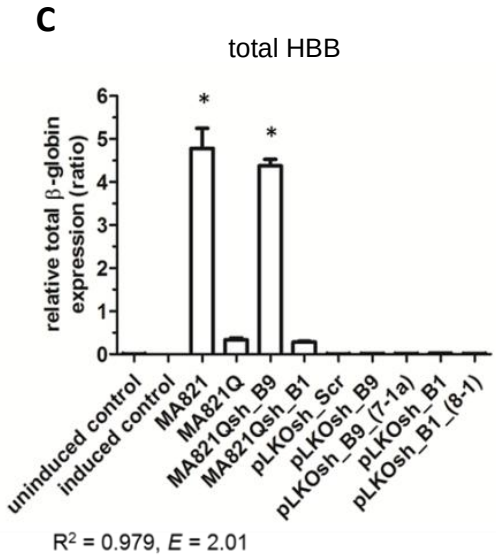
Moreover, the effect of BCL11A expression at the protein level measured by Western blot analysis was uninformative. Samples were investigated on three separate immunoblots with extracts from sets of BCL11A_451-shRNA (**Figure 3.3.6A**), BCL11A_449-shRNA (**Figure 3.3.6B**), and MA821Q-derived (**Figure 3.3.6C**) vectors. Expression of BCL11A protein was inconclusive as a result of a persistent background, white spots and random speckles on the blots. Expression of HBG protein was detected in all samples. Quantification of HBG protein levels in cultures transduced with the BCL11A_451-shRNA LVs was not available owing to a weak or absent signal for the loading control protein (**Figure 3.3.6A**). However, expression of HBG protein increased in cultures transduced with the BCL11A_449-shRNA LVs as compared with control cultures. This increase was not considered statistically significant due to the large variation in HBG protein levels among duplicate samples of each vector. Our results showed

that transductions with the U6-based pLKOsh_B9_(7-1a) and pLKOsh_B9 vectors led to an increase in HBG protein by 72.8±7.1% and 320.1±191.1% respectively. Unexpectedly, we also observed a similar increase in HBG protein of 199.1% in cells transduced with the pLKOsh_Scr control vector (**Figure 3.3.6B**). In addition, analysis of MA821Qsh_B9 HBB-based LV transduced cells was troubling since one transduction led to an increase in HBG by 66.3% whereas the other led to a decrease of 73.2%. Transduction of MA821Qsh_B1 led to a decrease in HBG protein by 73.9±13.1%, which was comparable to levels detected in cells transduced with the control non-shRNA containing MA821 and MA821Q vectors (**Figure 3.3.6C**).

As expected, expression of HBB mRNA and protein profoundly increased in cultures transduced with *HBB*-encoding MA821Q-derived vectors compared with those transduced with the U6-based pLKO.1-based LVs or the non-transduced control cultures. We observed higher *HBB* expression with an increasing number of integrated vector copies; e.g., transduction of MA821 at 117±18.4 average VCN per cell led to approximately 1330% higher HBB mRNA levels compared with transduction of MA821Q at a VCN of 11.0±1.5 (**Figure 3.3.4C**). Further investigation for possible competition between vector *HBB* and endogenous *HBG* for the transcriptional machinery was uninformative. Increased *HBB* expression in cultures transduced with MA821Q-derived vectors associated with a concomitant increase in HBG mRNA levels (**Figure 3.3.4A and 4C**), while expression analysis of HBG at the protein level did not corroborate mRNA data (**Figure 3.3.4C and 6C**). The data did not have sufficient statistical power to deduce significance with respect to the basal level of *HBG* expression in control cultures. Based on previous observations that the expression of *HBB* would plateau after vector integration at high average vector copy number as a result of limiting transcription factor accessibility [379], we speculated that using an MOI of 50 could result in a high VCN per genome such that the *HBB* transgene would outcompete endogenous *HBG* and biasing transcription towards *HBB* expression. Consequently, we performed a follow-up experiment with lower vector doses.



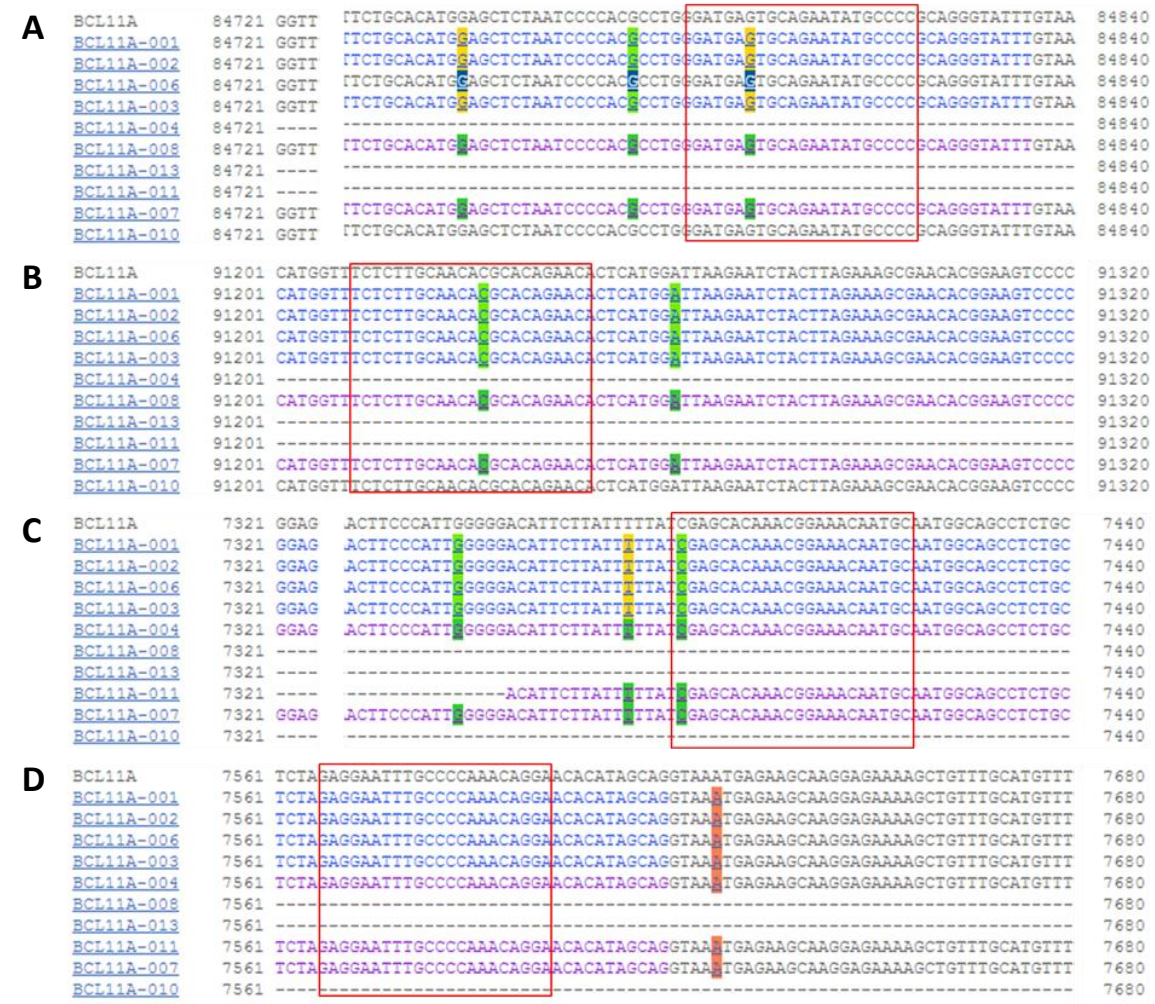




	relative gene expression						VCN (±SD)
	HBG		BCL11A_(B1)		total HBB		
	ratio (±SD)	Δ% (fold change)	ratio (±SD)	Δ% (fold change)	ratio (±SD)	Δ% (fold change)	
uninduced control	0.49±0.1		0.68±0.1		0.007±0.0		
induced control	0.55±0.1		0.18±0.1		0.004±0.0		
MA821	0.94±0.1	70.13 (1.70)	0.26±0.0	43.29 (1.43)	4.78±0.7	132273.85 (1323.74)	117.2±18.4
MA821Q	0.71±0.0	27.50 (1.27)	0.50±0.0	181.98 (2.82)	0.34±0.1	9165.91 (92.66)	11.0±1.5
MA821Qsh_B9	0.64±0.1	14.57 (1.15)	0.69±0.4	284.57 (3.85)	4.38±0.2	121073.61 (1211.74)	86.0±8.2
MA821Qsh_B1	0.58±0.1	5.08 (1.05)	0.39±0.1	117.83 (2.18)	0.28±0.0	7762.07 (78.62)	10.1±2.9
pLKOsh_Scr	0.56±0.1	1.28 (1.01)	0.43±0.4	140.10 (2.40)	0.006±0.0	56.86 (1.57)	11.5±1.1
pLKOsh_B9	2.01±0.2	262.74 (3.63)	0.49±0.1	178.57 (2.79)	0.010±0.0	179.01 (2.79)	72.0±14.4
pLKOsh_B9_(7-1a)	2.41±0.2	334.74 (4.35)	0.81±0.5	349.95 (4.50)	0.016±0.0	338.81 (4.39)	84.8±4.5
pLKOsh_B1	0.85±0.1	54.21 (1.54)	1.59±0.3	789.59 (8.90)	0.020±0.0	462.40 (5.62)	118.4±17.9
pLKOsh_B1_(8-1)	0.60±0.0	8.58 (1.09)	0.38±0.4	113.35 (2.13)	0.011±0.0	198.65 (2.99)	87.0±11.6

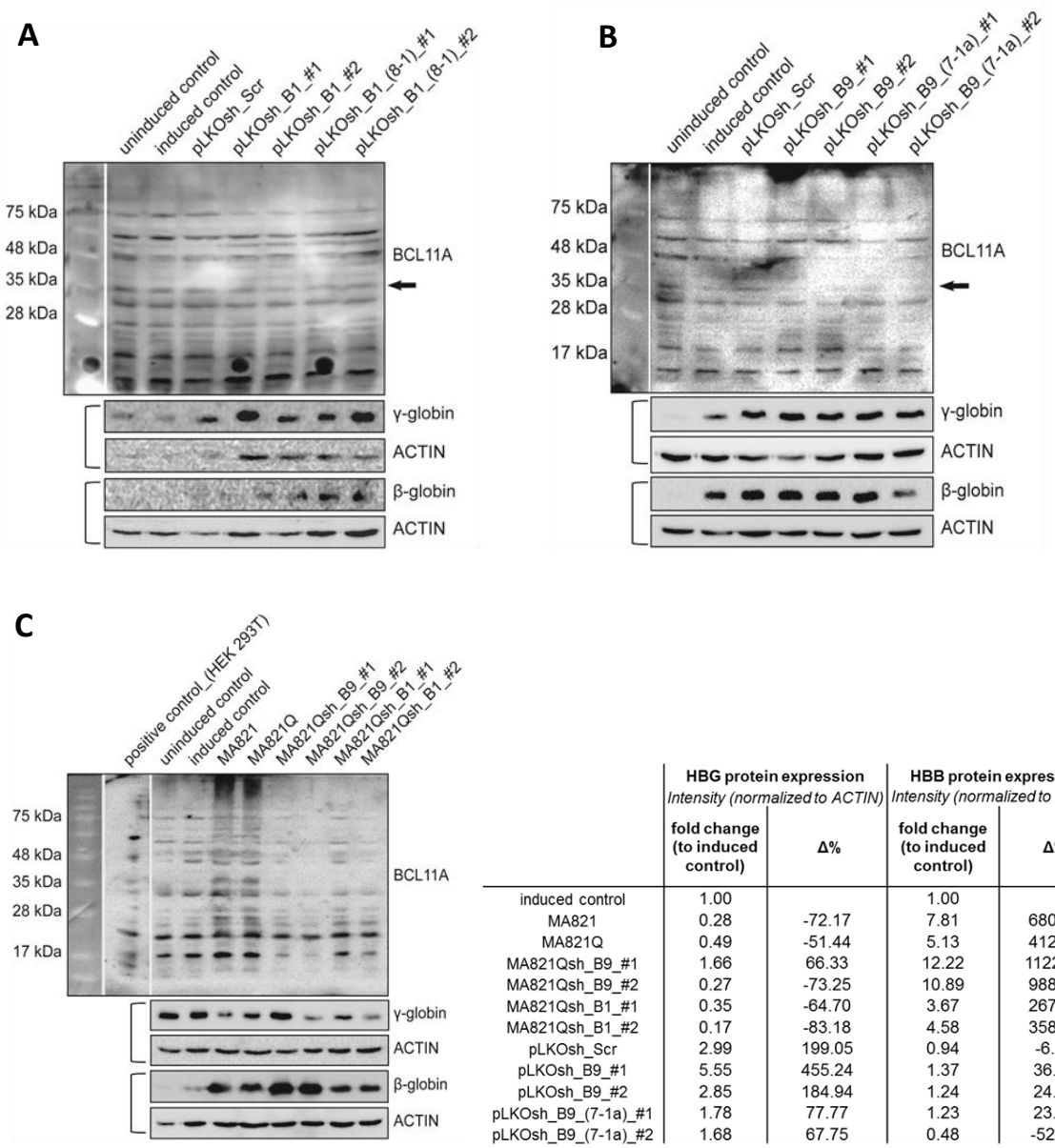
Figure 3.3.4. RT-qPCR analysis of mRNA expression in transduced K562 cells. K562 cells were transduced with various LVs as indicated at a MOI of 50, treated with 150 µM hydroxyurea, and analysed for **A)** HBG, **B)** BCL11A (region surrounding the B1 recognition site), and **C)** total HBB (endogenous and vector-derived) mRNA levels by RT-qPCR on day 7 post-transduction. Values are normalized by *GAPDH* expression. The histograms are mean ± SD from one experiment of duplicate sets. The table shows relative gene expression, the percentage change (Δ%) and fold change in expression between the transduced and the control (induced, untransduced) cultures, as well as the average cell VCN in bulk cell samples. Statistical analysis was performed by two-way ANOVA. Statistical different values (vs. induced control) are annotated by (*, $p<0.05$). E: PCR amplification efficiency; R^2: square of Pearson's correlation coefficient.

Name	Transcript ID	Biotype	RefSeq	bp	Protein
BCL11A-001	ENST00000335712	Protein coding	NM_022893; NP_075044	5942	835 aa
BCL11A-002	ENST00000356842	Protein coding	NM_018014; NP_060484	3959	773 aa
BCL11A-003	ENST00000359629	Protein coding	NM_138559; NP_612569	2427	243 aa
BCL11A-006	ENST00000358510	Protein coding	-	2697	801 aa
BCL11A-008	ENST00000477659	Processed transcript	-	2775	-
BCL11A-007	ENST00000489516	Processed transcript	-	1508	-
BCL11A-004	ENST00000409351	Processed transcript	-	659	-
BCL11A-013	ENST00000479026	Processed transcript	-	567	-
BCL11A-010	ENST00000492272	Processed transcript	-	562	-
BCL11A-011	ENST00000489183	Processed transcript	-	372	-



	primer	sequence (5' - 3') - sense strand
A	shB449_FW	GATGAGTGCAGAATATGCCCCG
B	shB449_RV	TCTCTTGCAACACGCACAGAACA
C	shB451_FW	CGAGCACAAACGGAAACAATGC
D	shB451_RV	GAGGAATTTGCCCCAAACAGGA

Figure 3.3.5. Alignment of BCL11A-specific primers against known human BCL11A transcript variants. Alignment confirms specificity of primers to target sequences (squared in red) in human BCL11A transcript variants annotated in *Ensembl*.



	HBG protein expression *Intensity (normalized to ACTIN)*		HBB protein expression *Intensity (normalized to ACTIN)*	
	fold change (to induced control)	**Δ%**	**fold change (to induced control)**	**Δ%**
induced control	1.00		1.00	
MA821	0.28	-72.17	7.81	680.79
MA821Q	0.49	-51.44	5.13	412.57
MA821Qsh_B9_#1	1.66	66.33	12.22	1122.17
MA821Qsh_B9_#2	0.27	-73.25	10.89	988.93
MA821Qsh_B1_#1	0.35	-64.70	3.67	267.23
MA821Qsh_B1_#2	0.17	-83.18	4.58	358.19
pLKOsh_Scr	2.99	199.05	0.94	-6.19
pLKOsh_B9_#1	5.55	455.24	1.37	36.95
pLKOsh_B9_#2	2.85	184.94	1.24	24.32
pLKOsh_B9_(7-1a)_#1	1.78	77.77	1.23	23.06
pLKOsh_B9_(7-1a)_#2	1.68	67.75	0.48	-52.36

Figure 3.3.6. Expression levels of BCL11A, HBG and HBB protein in K562 cells by Western blot analysis. K562 cells were transduced with various LVs as indicated at a MOI of 50, and cultured for 7 days in the presence of 150 μM hydroxyurea. Western immunoblots were probed with antibodies against BCL11A, HBG and HBB in whole cell extracts from $2x10^5$ cells. ACTIN was used as internal control. The table shows fold change relative to the control (induced, untransduced) culture, which is arbitrarily set at 1.00, and the percentage change (Δ%) in expression between the transduced and the control cultures. Molecular weight of BCL11A isoforms XL/L/M/S/XS: 125/100/51/35/25 kDa.

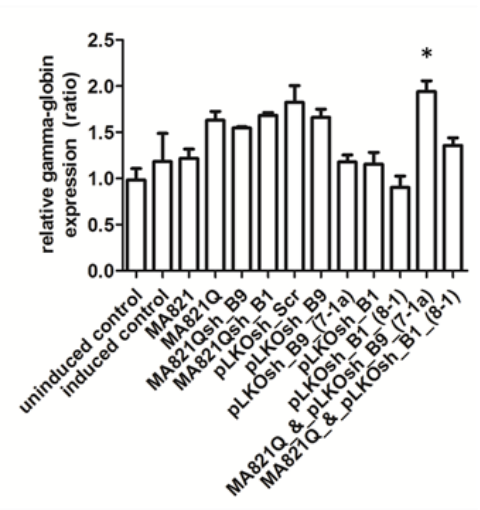
A large-scale experiment on all vectors was set-up with transductions performed in duplicate at low vector doses (MOI 10). Cultures co-transduced with the *HBB*-based MA821Q and the U6-based pLKO.1 BCL11A-shRNA vectors were included with the aim to test the functionality of the RNAi strategy from the MA821Qsh_BCL11A vectors with shRNAs targeting BCL11A inserted within the IVS-II of the *HBB* cassette. We observed a low degree of erythroid differentiation into haemoglobinized cells (10%–15%) for all cultures. In comparison to the induced non-transduced control culture (1.81±0.4), cultures treated with MA821Qsh_B9 and MA821Qsh_B1 LVs led to an increased expression of HBG mRNA by 30.9% (1.55±0.0) and 42.3% (1.68±0.1) respectively, which was comparable to HBG mRNA levels detected for the MA821 (1.22±0.1) and MA821Q (1.63±0.1) vector transduced cells. Transduction with U6-based pLKOsh_B9_(7-1a) produced comparable HBG expression (1.12±0.1) to that of the control whereas transduction with pLKOsh_B9 led to an increase in HBG mRNA of 40.6% (1.66±0.1). Surprisingly, transduction with the control pLKOsh_Scr vector led to a higher increase of 54.5% (1.82±0.3) in HBG mRNA. There was no detectable increase in the level of HBG expression for the pLKO.1 BCL11A_451-shRNA vectors (**Figure 3.3.7A**). As before, RT-qPCR analysis to measure BCL11A expression was uninformative; the shB9 primer failed to produce detectable amplicons while the shB1 primer set gave BCL11A mRNA levels comparable to those of control cultures (**Figure 3.3.7B**). Unexpectedly, co-transduction of cells with MA821Q and pLKOsh_B9_(7-1a) vectors led to significantly increased levels of HBG mRNA by 64.2% (1.94±0.2) with a concomitant increase in BCL11A mRNA expression (**Figure 3.3.7A and 7B**). Co-transductions were performed sequentially over a 15 hour time-period, at low vector doses (MOI 10) as singlet transductions, in order to reduce receptor interference for viral entry.

We showed that all cultures treated with *HBB*-based MA821Q-derived vectors led to increased expression of HBB mRNA (**Figure 3.3.7C**). Also, we observed that the co-transduced samples achieved an increased HBB mRNA without affecting expression of endogenous *HBG*, which was found to be at levels higher or comparable to those of the control culture (**Figure 3.3.7A**).

The effect on BCL11A expression at the protein level measured by Western blot analysis was ambiguous since duplicate sample sets of each vector transduction gave different signal target intensities. Samples were investigated on three separate immunoblots in sets of samples from

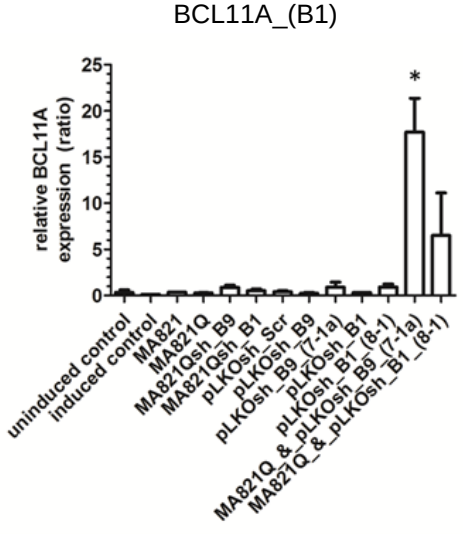
cells transduced with BCL11A_449-shRNA (**Figure 3.3.8A**), BCL11A_451-shRNA (**Figure 3.3.8B**), and MA821Q-derived (**Figure 3.3.8C**) vectors. Of note, protein expression on the two former blots was normalized to ACTIN whereas protein expression on this latter blot was normalized to GAPDH. Co-transduced samples were loaded and run on both types of blots. Our results showed that duplicate transductions with the U6-based pLKOsh_B9_(7-1a) vector led to an increase in HBG protein levels by 93.6% and 462.6%. Similarly, duplicate transductions with pLKOsh_B9 led to an increase in HBG protein levels by 45.5% and 232.2%. We did not observe an increase in the level of HBG expression in cells transduced with the pLKO.1 BCL11A_451-shRNA vector. Again quite unexpectedly, we showed that duplicate transductions with the control pLKOsh_Scr vector led to an increase in HBG protein levels by 69.9% and 140.9%. Also, none of the MA821Qsh_B9 or MA821Qsh_B1 vectors led to increased HBG expression at levels higher than those present in control vector transduced cells. In addition, we observed increased expression of HBB protein in all cultures transduced with the MA821Q-derived vectors. However, we found different expression levels of HBB and HBG protein in co-transduced samples run on separate Western blots and normalized to ACTIN or GAPDH. Co-transduction of cells with *HBB*-based MA821Q and U6-based pLKOsh_B9_(7-1a) vectors led to an increase in HBB of 5448.6% and 9504.9% with a concomitant fold increase in HBG of 195.5% and 152.5% respectively, following normalization to ACTIN (**Figure 3.3.8A**). However, this co-transduction led to an increase in HBB mRNA of 135.1% and 1413.6% with a concomitant increase in HBG of 118.3% and 17.2% respectively, following normalization to GAPDH (**Figure 3.3.8C**). Despite differences in the size of the normalized quantities, both analyses exhibited comparable protein expression levels and indicated co-expression of the vector-derived HBB and the endogenous HBG. Lastly, expression of BCL11A protein was inconclusive as a result of a persistent high background on the immunoblots.

A



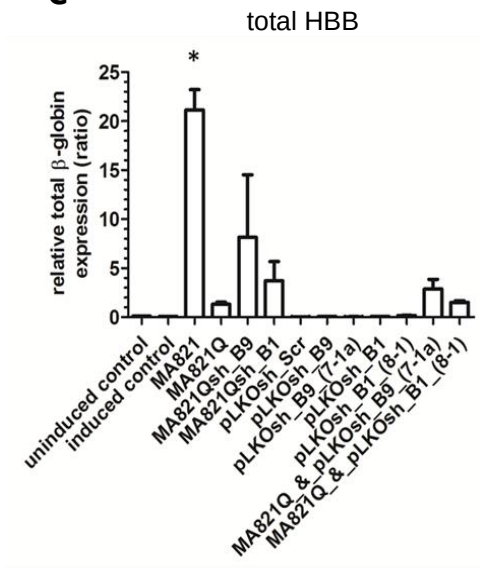
R^2 = 0.986, E = 2.27

B



R^2 = 0.848, E = 2.37

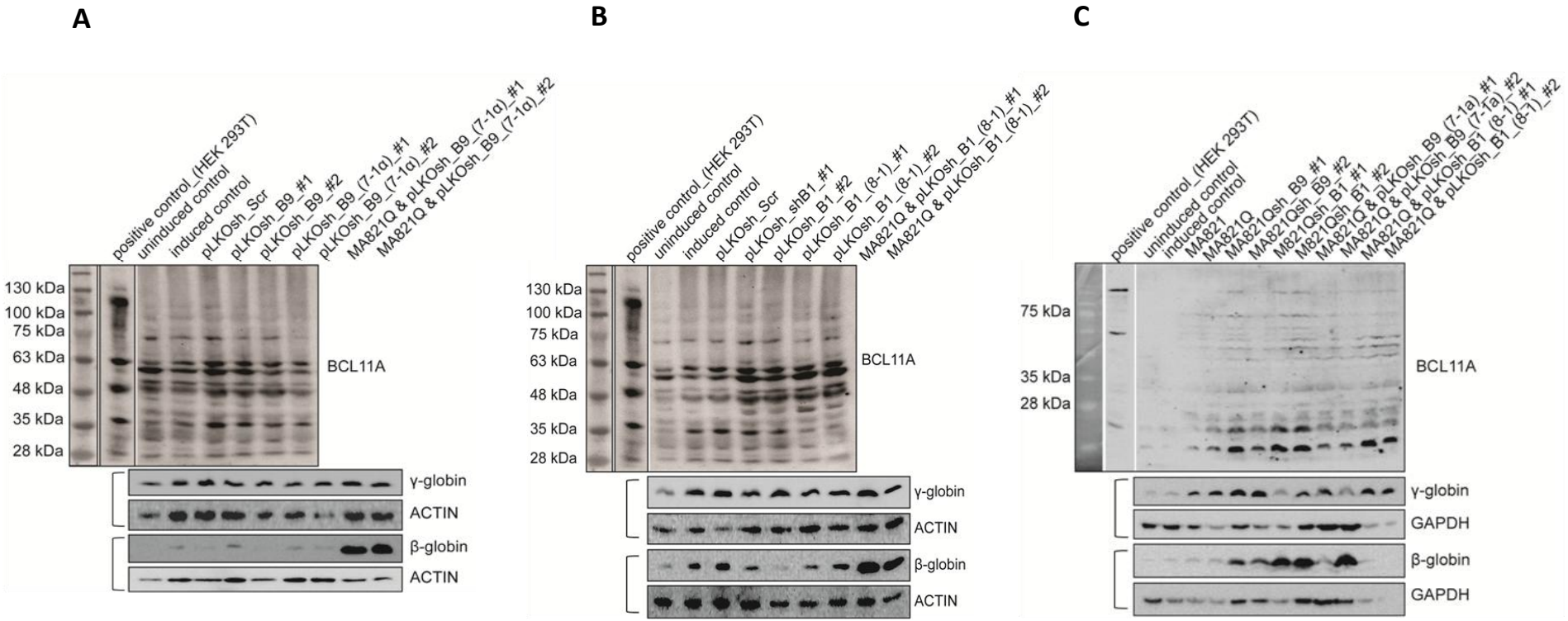
C



R^2 = 0.998, E = 1.86

	relative gene expression						VCN (±SD)
	HBG		BCL11A_(B1)		total HBB		
	ratio (±SD)	Δ% (fold change)	ratio (±SD)	Δ% (fold change)	ratio (±SD)	Δ% (fold change)	
uninduced control	0.98±0.2		0.35±0.4		0.11±0.0		N/A
induced control	1.81±0.4		0.14±N/A		0.09±N/A		N/A
MA821	1.22±0.1	3.00 (1.03)	0.37±N/A	164.94 (2.65)	21.16±2.3	23910.82 (240.11)	45.6±6.2
MA821Q	1.63±0.1	37.93 (1.38)	0.28±0.1	104.92 (2.05)	1.32±0.3	1401.56 (15.02)	27.8±1.9
MA821Qsh_B9	1.55±0.0	30.92 (1.31)	0.89±0.3	544.11 (6.44)	8.16±9.0	9165.10 (92.65)	37.2±8.7
MA821Qsh_B1	1.68±0.1	42.34 (1.42)	0.54±0.2	287.97 (3.88)	3.71±2.8	4114.87 (42.15)	23.8±6.4
pLKOsh_Scr	1.82±0.3	54.40 (1.54)	0.41±0.2	195.36 (2.95)	0.029±0.0	-67.26 (0.33)	8.3±1.2
pLKOsh_B9	1.66±0.1	40.61 (1.41)	0.23±0.1	79.12 (1.79)	0.074±0.0	-16.42 (0.84)	3.6±N/A
pLKOsh_B9_(7-1a)	1.12±0.1	-0.34 (1.00)	0.91±0.8	556.38 (6.56)	0.058±0.0	-34.60 (0.65)	2.6±0.6
pLKOsh_B1	1.15±0.2	-2.28 (0.98)	0.32±N/A	130.38 (2.30)	0.075±0.0	-14.93 (0.85)	11.2±2.4
pLKOsh_B1_(8-1)	0.90±0.2	-23.62 (0.76)	0.94±0.4	577.87 (6.78)	0.17±0.0	89.80 (1.90)	38.1±4.4
MA821Q_&_pLKOsh_B9_(7-1a)	1.94±0.2	64.19 (1.64)	17.71±5.2	12732.73 (128.33)	2.89±1.4	3177.46 (32.77)	52.9±34.8
MA821Q_&_pLKOsh_B1_(8-1)	1.36±0.1	14.81 (1.15)	6.50±6.5	4614.12 (47.14)	1.51±0.2	1614.48 (17.14)	N/A

Figure 3.3.7. RT-qPCR analysis of mRNA expression in K562 cells transduced at a MOI of 10. K562 cells were transduced with various LVs as indicated at a MOI of 10, treated with 150 μM hydroxyurea, and analysed for **A)** HBG, **B)** BCL11A (region surrounding the B1 recognition site), and **C)** total HBB (endogenous and vector-derived) mRNA levels by RT-qPCR on day 7 post-transduction. Values are normalized by *GAPDH* expression. The histograms are mean ± SD from one experiment of duplicate sets. The table shows the relative gene expression, the percentage change (Δ%) and fold change in expression between the transduced and the control (induced, untransduced) cultures, as well as as the average cell VCN in bulk cell samples. Statistical analysis was performed by two-way ANOVA. Statistical different values (vs. control) are annotated (*, p<0.05). E: PCR amplification efficiency; R^2: square of Pearson's correlation coefficient.


A
positive control_(HEK 293T)
uninduced control
induced control
pLKOsh_Scr
pLKOsh_B9_#1
pLKOsh_B9_#2
pLKOsh_B9_(7-1α)_#1
pLKOsh_B9_(7-1α)_#2
MA821Q & pLKOsh_B9_(7-1α)_#1
MA821Q & pLKOsh_B9_(7-1α)_#2
130 kDa
100 kDa
75 kDa
63 kDa
48 kDa
35 kDa
28 kDa
BCL11A
γ-globin
ACTIN
β-globin
ACTIN
B
positive control_(HEK 293T)
uninduced control
induced control
pLKOsh_Scr
pLKOsh_shB1_#1
pLKOsh_B1_#2
pLKOsh_B1_(8-1)_#1
pLKOsh_B1_(8-1)_#2
MA821Q & pLKOsh_B1_(8-1)_#1
MA821Q & pLKOsh_B1_(8-1)_#2
130 kDa
100 kDa
75 kDa
63 kDa
48 kDa
35 kDa
28 kDa
BCL11A
γ-globin
ACTIN
β-globin
ACTIN
C
positive control_(HEK 293T)
uninduced control
induced control
MA821
MA821Q
MA821Qsh_B9_#1
MA821Qsh_B9_#2
M821Qsh_B1_#1
MA821Qsh_B1_#2
MA821Q & pLKOsh_B9_(7-1α)_#1
MA821Q & pLKOsh_B9_(7-1α)_#2
MA821Q & pLKOsh_B1_(8-1)_#1
MA821Q & pLKOsh_B1_(8-1)_#2
75 kDa
35 kDa
28 kDa
BCL11A
γ-globin
GAPDH
β-globin
GAPDH

A–B

	HBG protein expression *Intensity (normalized to ACTIN)*		HBB protein expression *Intensity (normalized to ACTIN)*	
	fold change (to induced control)	**Δ%**	**fold change (to induced control)**	**Δ%**
induced control	1.00		1.00	
pLKOsh_Scr_#1	1.70	69.91	0.93	-7.12
pLKOsh_Scr_#2	2.41	140.93	2.03	102.70
pLKOsh_B9_#1	1.46	45.52	1.05	4.54
pLKOsh_B9_#2	3.32	232.22	0.64	-35.96
pLKOsh_B9_(7-1a)_#1	1.94	93.60	0.73	-27.36
pLKOsh_B9_(7-1a)_#2	5.63	462.57	0.63	-36.70
pLKOsh_B1_#1	0.47	-53.33	0.58	-42.03
pLKOsh_B1_#2	0.69	-31.27	0.00	0.00
pLKOsh_B1_(8-1)_#1	0.25	-74.67	1.11	10.62
pLKOsh_B1_(8-1)_#2	0.84	-15.55	3.53	252.58
MA821Q_&_pLKOsh_B9_(7-1a)_#1	2.96	195.53	55.49	5448.60
MA821Q_&_pLKOsh_B9_(7-1a)_#2	2.52	152.47	96.05	9504.87
MA821Q_&_pLKOsh_B1_(8-1)_#1	0.84	-16.26	59.92	5891.60
MA821Q_&_pLKOsh_B1_(8-1)_#2	0.57	-43.28	59.85	5884.53

C

	HBG protein expression *Intensity (normalized to GAPDH)*		HBB protein expression *Intensity (normalized to GAPDH)*	
	fold change (to induced control)	**Δ%**	**fold change (to induced control)**	**Δ%**
induced control	1.00		1.00	
MA821	3.64	264.07	1.99	98.92
MA821Q	10.09	908.67	2.87	187.22
MA821Qsh_B9_#1	10.53	953.01	6.83	583.43
MA821Qsh_B9_#2	20.21	1921.16	8.90	789.88
MA821Qsh_B1_#1	3.48	247.73	30.63	2963.18
MA821Qsh_B1_#2	2.66	165.76	10.60	960.28
MA821Q_&_pLKOsh_B9_(7-1a)_#1	2.18	118.31	2.35	135.09
MA821Q_&_pLKOsh_B9_(7-1a)_#2	1.17	17.16	15.14	1413.61
MA821Q_&_pLKOsh_B1_(8-1)_#1	20.87	1987.32	3.04	204.41
MA821Q_&_pLKOsh_B1_(8-1)_#2	44.26	4325.61		

Figure 3.3.8. Expression levels of BCL11A, HBG and HBB protein in K562 cells by Western blot analysis. K562 cells were transduced with various LVs as indicated at a MOI of 10 and cultured for 7 days in the presence of 150 μM hydroxyurea. Western immunoblot analysis of whole cell extracts from $2x10^5$ cells were probed with antibodies against BCL11A, HBG, and HBB. ACTIN was used as internal control. The table shows fold change relative to the control (induced, untransduced) sample, which is arbitrarily set at 1.00, and the percentage change (Δ%) in expression between the transduced and the control cultures. Molecular weight of BCL11A isoforms XL/L/M/S/XS: 125/100/51/35/25 kDa.

3.3.3 Characterization of BCL11A-targeting LVs in HEL cells

The methodology established for K562 cells was re-employed in the investigation of vector function and BCL11A knockdown in HEL cells. Despite their failure as a GFP reporter line (**section 3.2.3.3**), HEL cells have turned out to be a suitable platform for vector characterization since they have been shown to express the full-length BCL11A isoform that is required to modulate Hb switching [119]. In order to improve the specific BCL11A protein signal and reduce background bands on Western blots, we compared the use of three primary antibodies targeting this molecule in this analysis, namely from *Santa Cruz*, *Abcam* and *Benthyl*. These antibodies were evaluated for their ability to detect full-length BCL11A-XL (~120 kDa) and BCL11A-L (~100 kDa) in whole protein lysates from $2x10^5$ HEL cells (**Figure 3.3.9**). We showed that the *Abcam* antibody performed poorly, whereas the *Benthyl* and *Santa Cruz* antibodies produced clear signals for the large BCL11A isoforms in spite of a persistent background. In addition, clear areas and random speckles were occasionally detected on blots probed with any of the three anti-BCL11A antibodies. Therefore, preceded experiments were performed with the *Santa Cruz* antibody which was also validated by the manufacturer in lysates of HEL cells to detect the large BCL11A isoforms.

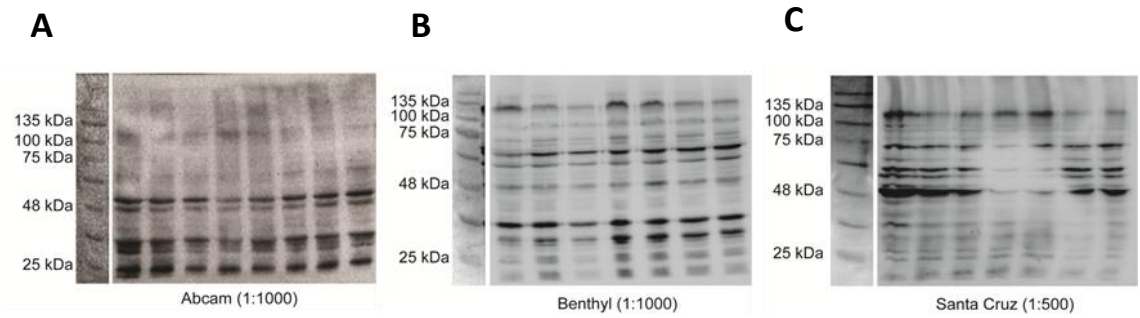


Figure 3.3.9. Expression levels of BCL11A protein in whole-cell extracts from HEL cells by Western blot analysis: comparison of three different commercial preparations of anti-BCL11A antibodies. The expression of BCL11A protein was quantified in lysates of $2x10^5$ HEL cells, which were loaded and run on blots probed with primary antibodies from **A**) *Abcam* (1:1000 dilution), **B**) *Benthyl* (1:1000 dilution), and **C**) *Santa Cruz* (1:500 dilution). Molecular weight of BCL11A isoforms XL/L/M/S/XS: 125/100/51/35/25 kDa.

Initially, a large-scale experiment on all vectors was set-up with transductions performed in singlet at a high vector dose (MOI 50) in order to maximise the possibility of detecting even a modest degree of BCL11A knockdown by Western blot analysis as well as to assess the specificity of the RT-qPCR BCL11A-primers for their target sequence in HEL cells. Cultures co-transduced with the *HBB*-based MA821Q and the U6-based pLKO.1 BCL11A-shRNA vectors were included with the aim to test the functionality of the RNAi strategy from the MA821Qsh_BCL11A vectors where the shRNA is located within IVS-II of the *HBB*. Previous reports showed that co-transduction of HEL cells with vectors targeting the same amphotropic receptor for cell entry resulted in reduced transduction efficiency [380]. Using vector doses at which the likelihood of receptor interference was minimum resulted in efficient and less random secondary transductions [381]. Therefore, we performed co-transductions with both LVs simultaneously at a MOI of 50 (similar to singlet transductions) and at a lower MOI of 25 in order to account for possible differences in vector transducibility, integration and expression. In this series of experiments, induced erythroid differentiation was performed with 50 μM haemin. The proportion of differentiated cells was determined by benzidine staining at the end of a 7-day period. The expression of BCL11A, HBG and HBB mRNA and protein levels was assessed by RT-qPCR and Western blotting respectively.

The fraction of benzidine-positive cells in the control culture (induced, untransduced) was ~6 times higher compared to the undifferentiated culture (40.5% vs. 6.5%), indicating that HEL cells in active growth had low levels of spontaneous differentiation. Transductions led to a similar or higher fraction of benzidine-positive cells compared to the control, and peaked at 66.5% for the U6-based pLKOsh_B9_(7-1a) vector (**Figure 3.3.10**). The experimental procedure led to cell death in the range of 10%–30%, possibly due to the result of metabolic acidosis caused by active cell growth together with transductions at high vector doses (**Figure 3.3.10**). We observed that haemin induction led to an increase in HBG mRNA of ~110% compared with basal HBG levels in undifferentiated cells (1.10 vs. 0.53). In addition, we found that none of the shRNA-expressing vectors led to a detectable increase in HBG expression at levels higher than those of the control vectors. Cultures transduced with MA821Qsh_B9 and MA821Qsh_B1 vectors led to an increase in HBG mRNA by 39.04% (1.53) and 0.08% (1.10) respectively

compared with the control culture (1.10), while similar levels of HBG were detected with the control MA821 (1.07), MA821Q (1.54), and MA821Qsh_G (1.51) vectors. Cultures transduced with the U6-based pLKOsh_B9_(7-1a) (1.03) and pLKOsh_B1_(8-1) (0.62) vectors produced similar or lower levels of HBG mRNA to that we saw in cells with the control vector pLKOsh_Scr (1.03) (**Figure 3.3.11A**). Quantification of BCL11A expression was uninformative since different RT-qPCR assays using primers amplifying different BCL11A siRNA target sequences (shB9 and shB1) gave variable readings for cultures that were not treated with shRNA-expressing vectors (**Figure 3.3.11C and 11D**). Also, we observed that increase in *HBG* expression was not always associated with a concomitant reduction in BCL11A. Importantly, the detection of amplicons using either primer set indicated that the shRNA-targeted sequences were present in the transcriptome of HEL cells. The absence of detectable BCL11A knockdown using large vector doses with the pLKO.1 BCL11A-shRNA vectors raised concerns over the functionality of the BCL11A-shRNA design. As the BCL11A-specific primers were designed to target regions common to the large and short BCL11A transcript isoforms, we next interrogated the activity of the BCL11A-shRNAs on the expression levels of the large isoforms through Western blotting.

The effect of BCL11A expression at the protein level corresponded to findings at the mRNA level. We showed that none of the shRNA-expressing vectors led to a detectable increase in *HBG* expression at levels higher than those seen with control vectors while knockdown of the BCL11A-XL was concomitant with either an increase or a decrease in HBG chain synthesis. Analysis of U6-based pLKO.1 BCL11A-shRNA vectors, which would inform on the functionality of the design of the shRNA sequence itself, was inconclusive. Our results showed that transduction with pLKOsh_B9_(7-1a) led to an increase in HBG protein by 39.2% and a decrease in BCL11A-XL protein by 60.6%. By comparison, transduction with pLKOsh_B1_(8-1) led to a decrease in both HBG and BCL11A-XL proteins by 16.7% and 72.9% respectively. Unexpectedly, transduction with the control pLKOsh_Scr vector also led to a higher increase in HBG protein by 48.4% and a decrease in BCL11A-XL by 22.5%. Furthermore, the sample cells co-transduced with non-shRNA containing MA821Q and the pLKOsh_B9_(7-1a) shRNA vectors at a MOI of 50 and 25 respectively, showed an increase in HBG protein of 201.1% and 132% respectively without a concomitant reduction in BCL11A-XL protein. Surprisingly, we were

unable to detect HBG protein expression in the co-transduced sample with MA821Q and pLKOsh_B1_(8-1) vectors (**Figure 3.3.12**). Of note, the co-transduced samples exhibited low levels of HBG mRNA expression comparable to those in the control cultures (**Figure 3.3.11A**). Although the Western blot findings provided evidence to suggest functionality for the BCL11A_449-shRNA design, when combined with the mRNA results, the overall data do not provide concrete support for this conclusion.

Expression of HBB mRNA was detected in all samples and, as expected, showed a profound increase in cultures treated with MA821Q-derived vectors, indicating that the cellular transcription machinery supported expression of the *HBB* transgene (**Figure 3.3.11B**). Towards the assessment of a possible competition between vector-based *HBB* and endogenous *HBG* expression, analysis of the co-transduced samples proved uninformative. We observed that the co-transduced samples, which received the MA821Q vector at a similar dose to the single transduction (MOI 50), had ~85% lower levels of HBB mRNA, suggesting inefficient transducibility. Since cultures received both vectors simultaneously, it could be argued that transduction efficiency with the U6-based pLKO.1 BCL11A-shRNA vector was similarly reduced. Moreover, quantification of HBB protein expression on the Western blots was prevented either by a weak signal intensity due to naturally low levels of HBB or as a result of inadequate amounts of protein sample loaded (**Figure 3.3.12**). Given the fact that HBG mRNA and protein levels were different, together with an incomplete set of data for *HBB* expression, the co-transduced samples were rendered uninformative and excluded from analysis for vector function.

In order to produce data with statistical power to deduce significance as well as to minimize the possibility for competition between the vector-based *HBB* and endogenous *HBG*, two subsequent experiments were set-up with transductions performed in duplicate to give an estimated VCN of 5. Unfortunately, one of these two experiments was excluded because of an average VCN ≤1 for all samples.

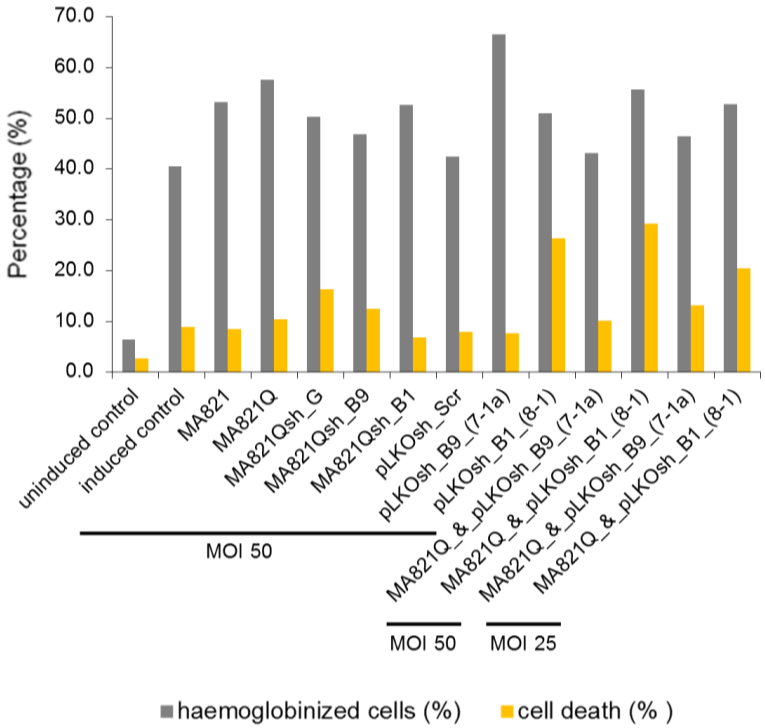
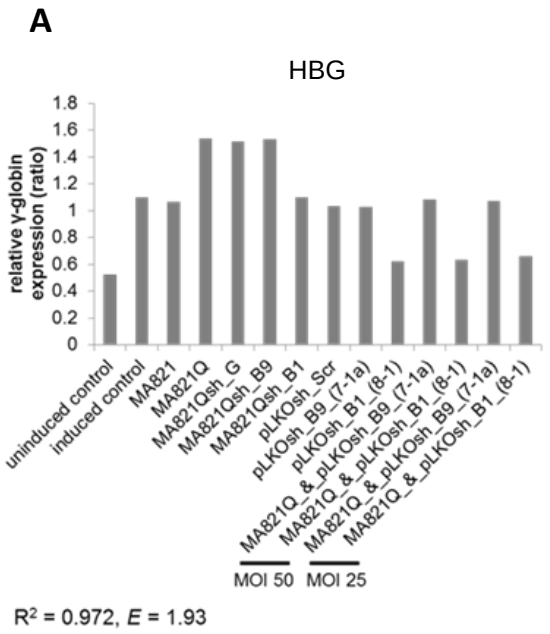
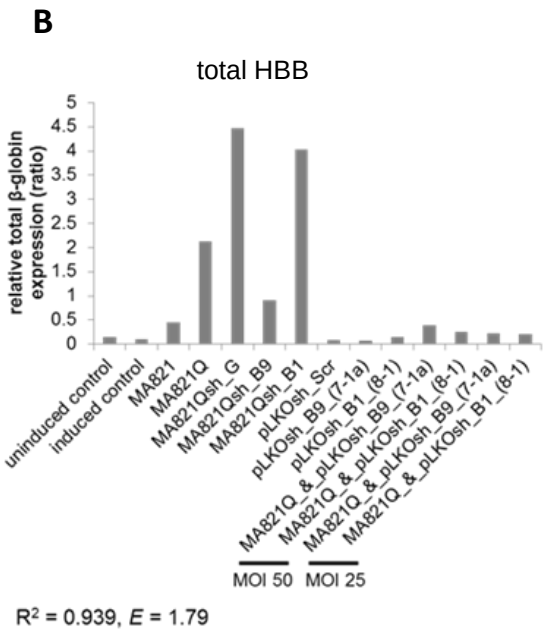


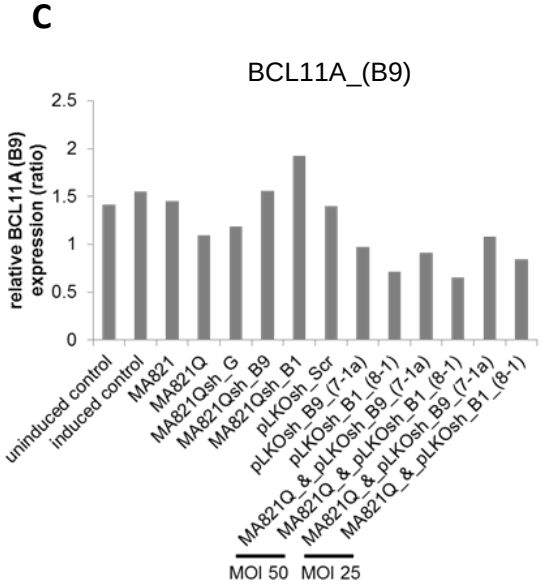
Figure 3.3.10. Erythroid induction of HEL cells. HEL cells were transduced with various LVs as indicated at a MOI of 25 or 50, cultured for 7 days in the presence of 50 μM haemin, and analysed for the percentage (%) of haemoglobinization (grey bars) and cell death (yellow bars) by benzidine and trypan blue staining respectively. Data are presented for single transductions from one experiment.



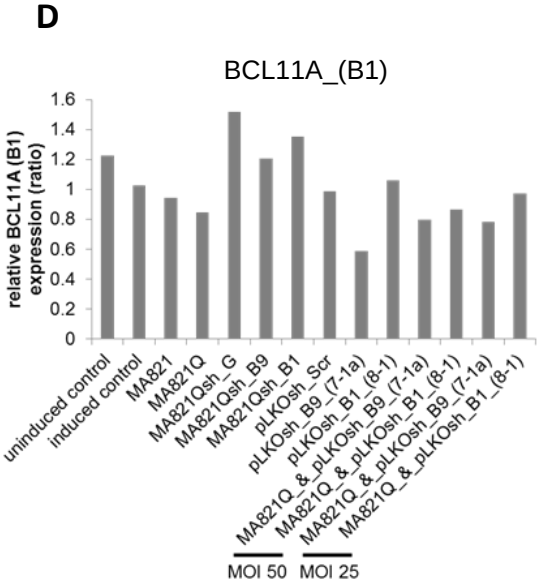
R^2 = 0.972, E = 1.93



R^2 = 0.939, E = 1.79



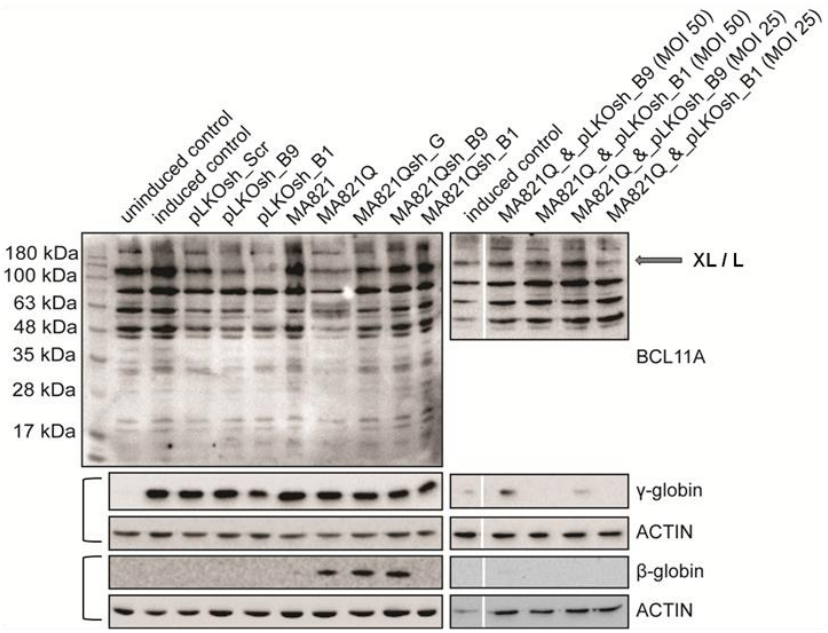
R^2 = 0.934, E = 1.65



R^2 = 0.913, E = 1.90

		relative gene expression								VCN
		HBG		BCL11A_(B9)		BCL11A_(B1)		total HBB		
		ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	
	uninduced control	0.53		1.41		1.23		0.15		
	induced control	1.10		1.55		1.03		0.10		
	MA821	1.07	-3.29 (0.97)	1.45	-6.01 (0.94)	0.94	-8.19 (0.92)	0.46	339.98 (4.40)	5.06
	MA821Q	1.54	39.81 (1.40)	1.10	-29.09 (0.71)	0.85	-17.74 (0.82)	2.13	1943.97 (20.44)	55.19
	MA821Qsh_G	1.51	37.50 (1.38)	1.19	-23.40 (0.77)	1.52	47.65 (1.48)	4.48	4196.31 (42.96)	35.97
	MA821Qsh_B9	1.53	39.04 (1.39)	1.56	0.91 (1.01)	1.21	17.62 (1.18)	0.91	770.80 (8.71)	30.85
	MA821Qsh_B1	1.10	0.08 (1.00)	1.93	24.38 (1.24)	1.35	31.62 (1.32)	4.04	3776.60 (38.77)	32.52
	pLKOsh_Scr	1.03	-6.25 (0.94)	1.40	-9.48 (0.91)	0.99	-4.00 (0.96)	0.08	-22.74 (0.77)	6.23
	pLKOsh_B9_(7-1a)	1.03	-6.60 (0.93)	0.97	-37.02 (0.63)	0.59	-42.62 (0.57)	0.07	-31.91 (0.68)	8.53
	pLKOsh_B1_(8-1)	0.62	-43.45 (0.57)	0.71	-54.04 (0.46)	1.06	3.05 (1.03)	0.15	42.22 (1.42)	56.59
MOI 50	MA821Q_&_pLKOsh_B9_(7-1a)	1.08	-1.84 (0.98)	0.91	-40.98 (0.59)	0.80	-22.47 (0.78)	0.40	280.74 (3.81)	
MOI 50	MA821Q_&_pLKOsh_B1_(8-1)	0.63	-42.37 (0.58)	0.65	-57.73 (0.42)	0.87	-15.58 (0.84)	0.25	138.13 (2.38)	
MOI 25	MA821Q_&_pLKOsh_B9_(7-1a)	1.07	-2.66 (0.97)	1.08	-30.42 (0.70)	0.78	-23.87 (0.76)	0.22	111.37 (2.11)	
MOI 25	MA821Q_&_pLKOsh_B1_(8-1)	0.66	-39.84 (0.60)	0.84	-45.50 (0.55)	0.97	-5.20 (0.95)	0.21	102.22 (2.02)	

Figure 3.3.11. RT-qPCR analysis of mRNA expression in HEL cells transduced at a MOI of 50. HEL cells were transduced with various LVs as indicated at a MOI of 25 or 50, treated with 50 µM haemin, and analysed for **A)** HBG, **B)** total HBB (endogenous and vector-derived), and **C–D)** BCL11A (regions surrounding the B9 and B1 recognition sites) mRNA levels by RT-qPCR on day 7 post-transduction. Values are normalized by *GAPDH* expression. Data are presented for single transductions from one experiment. The table shows relative gene expression, the percentage change (Δ%) and fold change in expression between the transduced and the control (induced, untransduced) cultures, as well as the average cell VCN in bulk cell samples. Statistical analysis was performed by two-way ANOVA. Statistical different values (vs. control) are annotated (*, $p<0.05$).



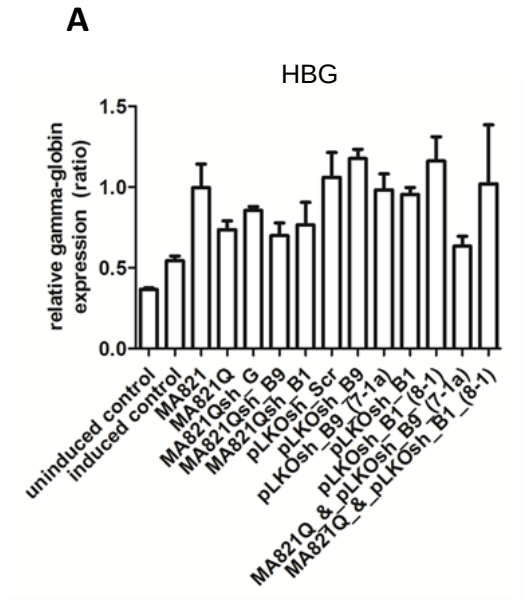
		HBG protein expression			BCL11A protein expression		
		Intensity (normalized to ACTIN)	fold change (to induced control)	Δ%	Intensity (normalized to ACTIN)	fold change (to induced control)	Δ%
	uninduced control	0.00			1.51		
	induced control	1.87	1.00		1.16	1.00	
	MA821	3.38	1.81	80.51	1.45	1.25	24.53
	MA821Q	3.38	1.80	80.45	0.79	0.68	-32.22
	MA821Qsh_G	2.28	1.22	21.77	0.76	0.65	-35.15
	MA821Qsh_B9	1.36	0.73	-27.19	0.83	0.71	-28.87
	MA821Qsh_B1	2.11	1.13	12.56	1.21	1.04	3.93
	pLKOsh_Scr	2.78	1.48	48.41	0.90	0.78	-22.49
	pLKOsh_B9_(7-1a)	2.61	1.39	39.24	0.46	0.39	-60.65
	pLKOsh_B1_(8-1)	1.56	0.83	-16.72	0.32	0.27	-72.96
	induced control	0.25	1.00		0.88	1.00	
MOI 50	MA821Q_&_pLKOsh_B9_(7-1a)	0.76	3.01	201.06	0.97	1.11	11.04
	MA821Q_&_pLKOsh_B1_(8-1)	N/A			1.00	1.14	14.35
MOI 25	MA821Q_&_pLKOsh_B9_(7-1a)	0.58	2.32	132.03	0.98	1.12	12.06
	MA821Q_&_pLKOsh_B1_(8-1)	0.12	0.48	-51.62	0.37	0.42	-58.27

Figure 3.3.12. Expression levels of BCL11A, HBG and HBB protein in HEL cells by Western blot analysis. HEL cells were transduced with various LVs as indicated at a MOI of 25 or 50, and cultured for 7 days in the presence of 50 μM haemin. Western immunoblots were probed with antibodies against BCL11A (*Santa Cruz*), HBG, and HBB in whole cell extracts from $2x10^5$ cells. ACTIN was used as internal control. The table shows expression ratio, fold change relative to the control (induced, untransduced) sample, which is arbitrarily set at 1.00, and the percentage change (Δ%) in expression between the transduced and the control cultures. Molecular weight of BCL11A isoforms XL/L/M/S/XS: 125/100/51/35/25 kDa.

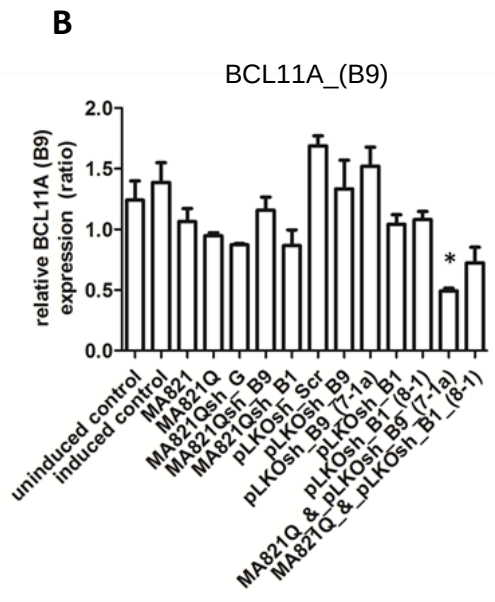
Analyses for the 2nd experiment were executed as described before. The fraction of benzidine-positive cells in the transduced cultures was comparable to the control culture (induced/untransduced) in the range of 35%–45% while spontaneous differentiation was scored at 7.9%. As before, haemin induction led to an increase in HBG mRNA levels of ~48.5% compared with basal HBG expression in undifferentiated cells (0.54±0.04 vs. 0.37±0.01) (**Figure 3.3.13A**). In addition, our results showed that none of the shRNA-expressing vectors led to significantly higher HBG mRNA levels in comparison to control cultures (0.54±0.04). Transductions of shRNA-containing *HBB*-based MA821Qsh_B9, and U6-based pLKOsh_B9_(7-1a) and pLKOsh_B9 vectors led to an increase in HBG mRNA of 28.4% (0.70±0.11), 80.5% (0.98±0.14), and 116.2% (1.18±0.08) respectively. Transductions with MA821Qsh_B1, pLKOsh_B1_(8-1), and pLKOsh_B1 gave an increase in HBG mRNA of 40.5% (0.77±0.20), 113.4% (1.16±0.21), and 75.2% (0.95±0.06) respectively. As we observed before, transductions with the control MA821Qsh_G and pLKOsh_Scr LVs also led to an increase in HBG mRNA of 57.2% (0.86±0.03) and 94.9% (1.06±0.22) respectively (**Figure 3.3.13A**). The lack of statistically significance differences between these test and control vectors indicated that the increase in *HBG* expression observed with the pLKO.1 BCL11A-shRNA vectors was due to random variation. In addition, knockdown of BCL11A expression was detected with primers amplifying different BCL11A siRNA target sequences in cultures treated with shRNA-expressing vectors as well as their controls (**Figure 3.3.13B and 13C**). Furthermore, expression levels of HBB mRNA were detected in all samples and again as expected showed a major increase in cultures transduced with MA821Q-derived vectors. Strikingly, compared to induced untransduced control (0.26±0.01) we found that transduction with MA821Qsh_B1 led to a low fold increase in HBB mRNA of 4.8% (0.27±0.06) with an average VCN per cell of 4.03±0.09, whereas, transduction of MA821Qsh_G at a similar VCN of 3.94±0.43 led to a significant increase in HBB mRNA of 717.9% (2.10±0.31) (**Figure 3.3.13D**). Also, we observed that transduction with the prototypical *HBB*-based MA821 vector led to an ncrease in HBB mRNA of 1702.2% (4.62±1.25) at an average VCN per cell of 7.47±0.42, which was significantly greater to the levels detected in cells transduced with MA821Q (1.28±0.48) at a similar VCN of 9.68±4.94 (**Figure 3.3.13D**).

Co-transductions were performed sequentially over a 15 hour time-period at low vector doses in order to reduce receptor interference for viral entry. The co-transduced samples received the MA821Q vector last at a similar dose to the single transduction (VCN 5) and resulted in ~76% lower *HBB* expression (0.48, 0.52 vs. 1.28±0.48) at levels comparable to those of the control vectors (0.4–0.5), indicating reduced transducibility (**Figure 3.3.13D**). As before, the co-transduced samples were uninformative for assessing vector functionality. The *HBG* expression data did not corroborate findings from the previous experiment indicating a functional BCL11A_449-shRNA design.

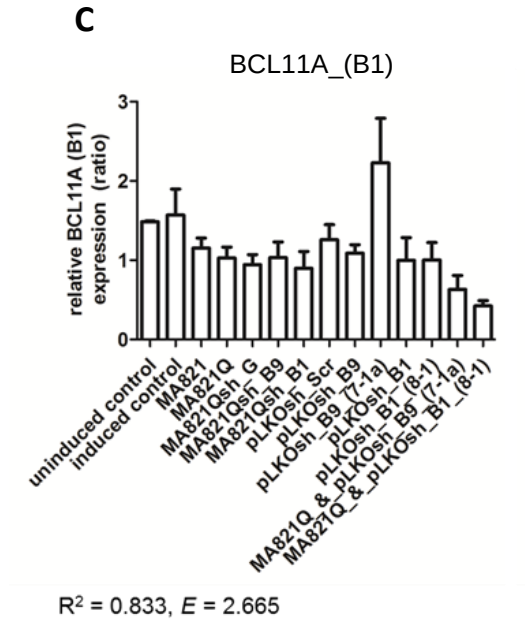
We then performed quantification of protein expression by Western blot analysis (**Figure 3.3.14**). Similar to the mRNA data, we observed that none of the BCL11A shRNA-expressing vectors led to an increase in *HBG* expression levels higher than those seen with control vectors. Unexpectedly, immunoblots of duplicate samples from a given vector sample run and analysed on separate gels, detected different amounts of HBG; e.g., transductions of MA821Q revealed a decrease in HBG protein of 5.9% on the 1st blot (**Figure 3.3.14A**) and an increase of 71.1% on the 2nd blot (**Figure 3.3.14B**), the latter representing the highest detectable signal for HBG protein expression among transduced samples. Further quantification of BCL11A protein was uninformative due to the presence of persistent high background on the immunoblots. In addition, HBB expression at the protein level was comparable to findings at the mRNA level, supporting findings of reduced *HBB* expression from the MA821Qsh_B1 shRNA-containing vector design. Also, Western blot data did not replicate a detectable difference in HBB mRNA levels between the MA821 and MA821Q vectors at a comparable VCN. Findings on the 1st blot revealed an increase in HBB by 1152.2% and 2040.1% respectively, whereas, findings on the 2nd blot revealed an opposite effect with an increase in HBB of 5694.1% and 1470.6% respectively between the two vectors. Moreover, the discrepancy in protein expression between duplicate samples of these vectors on separate blots provided evidence for and against possible competition between vector-derived *HBB* and endogenous *HBG* for the transcriptional machinery. Consequently, these inconsistencies rendered the Western blot data unreliable for assessing vector function while also reducing credibility for the mRNA expression data.



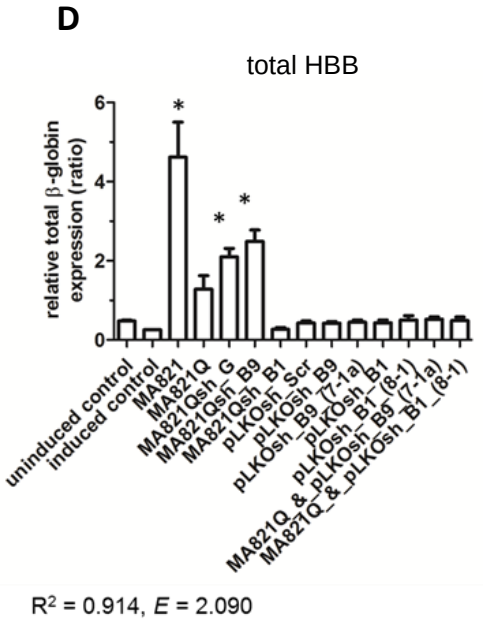
$R^2 = 0.981$, $E = 2.099$



$R^2 = 0.939$, $E = 2.064$



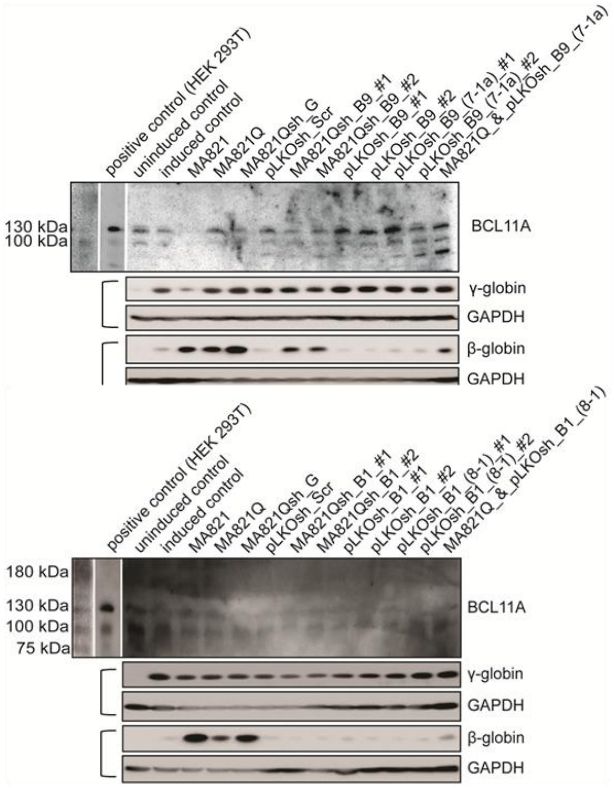
$R^2 = 0.833$, $E = 2.665$



$R^2 = 0.914$, $E = 2.090$

	relative gene expression								VCN (±SD)
	HBG		BCL11A_(B9)		BCL11A_(B1)		total HBB		
	ratio (±SD)	Δ% (fold change)	ratio (±SD)	Δ% (fold change)	ratio (±SD)	Δ% (fold change)	ratio (±SD)	Δ% (fold change)	
uninduced control	0.37±0.01		1.24±0.23		1.49±0.02		0.48±0.03		
induced control	0.54±0.04		1.39±0.23		1.57±0.46		0.26±0.01		
MA821	0.997±0.21	83.05 (1.83)	1.06±0.15	-23.34 (0.77)	1.15±0.18	-26.48 (0.74)	4.62±1.25	1702.15 (18.02)	7.47±0.42
MA821Q	0.74±0.08	35.29 (1.35)	0.95±0.04	-31.72 (0.68)	1.03±0.20	-34.57 (0.65)	1.28±0.48	401.43 (5.01)	9.68±4.94
MA821Qsh_G	0.86±0.03	57.23 (1.57)	0.87±0.02	-37.14 (0.63)	0.94±0.18	-39.97 (0.60)	2.10±0.31	717.87 (8.18)	3.94±0.43
MA821Qsh_B9	0.70±0.11	28.42 (1.28)	1.16±0.15	-16.55 (0.83)	1.03±0.28	-34.16 (0.66)	2.49±0.41	871.02 (9.71)	23.11±2.90
MA821Qsh_B1	0.77±0.20	40.53 (1.41)	0.87±0.18	-37.62 (0.62)	0.899±0.30	-42.74 (0.57)	0.27±0.06	4.78 (1.05)	4.03±0.09
pLKOsh_Scr	1.06±0.22	94.85 (1.95)	1.69±0.12	21.73 (1.22)	1.26±0.27	-19.86 (0.80)	0.43±0.07	65.90 (1.66)	6.31±7.97
pLKOsh_B9	1.18±0.08	116.16 (2.16)	1.33±0.34	-3.83 (0.96)	1.09±0.15	-30.72 (0.69)	0.42±0.05	65.74 (1.66)	1.21±0.22
pLKOsh_B9_(7-1a)	0.98±0.14	80.46 (1.80)	1.52±0.22	9.60 (1.10)	2.23±0.80	41.85 (1.42)	0.45±0.08	74.52 (1.75)	0.97±0.26
pLKOsh_B1	0.95±0.06	75.21 (1.75)	1.04±0.11	-24.96 (0.75)	0.998±0.41	-36.46 (0.64)	0.43±0.10	67.94 (1.68)	4.07±0.11
pLKOsh_B1_(8-1)	1.16±0.21	113.38 (2.13)	1.08±0.09	-22.03 (0.78)	1.00±0.31	-36.12 (0.64)	0.50±0.16	96.52 (1.97)	2.24±0.29
MA821Q_&_pLKOsh_B9_(7-1a)	0.63±0.09	16.42 (1.16)	0.49±0.03	-64.47 (0.36)	0.63±0.25	-59.77 (0.40)	0.52±0.09	103.00 (2.03)	7.77±4.43
MA821Q_&_pLKOsh_B1_(8-1)	1.02±0.52	87.54 (1.88)	0.72±0.18	-47.78 (0.52)	0.42±0.09	-73.01 (0.27)	0.49±0.13	89.78 (1.90)	7.21±2.14

Figure 3.3.13. RT-qPCR analysis of mRNA expression in transduced HEL cells. HEL cells were transduced with various LVs as indicated to give an estimated VCN of 5, treated with 50 μM haemin, and analysed for **A)** HBG, **B–C)** BCL11A (regions surrounding the B9 and B1 recognition sites), and **D)** total HBB (endogenous and vector-derived) mRNA levels by RT-qPCR on day 7 post-transduction. Values are normalized by *GAPDH* expression. The histograms are mean ± SD from one experiment of duplicate sets. The table shows relative gene expression, the percentage change (Δ%) and fold change in expression between the transduced and the control (induced, untransduced) cultures, as well as the average cell VCN in bulk cell samples. Statistical analysis was performed by two-way ANOVA. Statistically different values (vs. control) are annotated (*, $p<0.05$).



	HBG protein expression			HBB protein expression		
	Intensity (normalized to GAPDH)	fold change (to induced control)	Δ%	Intensity (normalized to GAPDH)	fold change (to induced control)	Δ%
uninduced control_#1	0.02			0.00		
induced control_#1	0.71	1.00		0.08	1.00	
MA821_#1	0.26	0.37	-62.92	0.95	12.52	1152.21
MA821Q_#1	0.67	0.94	-5.92	1.62	21.40	2040.14
MA821Qsh_G_#1	0.85	1.19	19.47	3.43	45.40	4440.17
MA821Qsh_B9_#1	0.92	1.30	29.94	1.39	18.37	1736.89
MA821Qsh_B9_#2	0.69	0.97	-2.66	1.26	16.63	1562.95
pLKOsh_Scr_#1	0.86	1.20	20.29	0.23	3.03	202.70
pLKOsh_B9_#1	1.04	1.46	45.77	0.12	1.63	62.62
pLKOsh_B9_#2	1.09	1.53	52.72	0.07	0.94	-6.24
pLKOsh_B9_(7-1a)_#1	0.87	1.22	21.84	0.14	1.87	86.77
pLKOsh_B9_(7-1a)_#2	0.67	0.94	-6.25	0.13	1.71	71.18
MA821Q_&_pLKOsh_B9_(7-1a)	0.70	0.99	-1.48	0.33	4.40	340.36

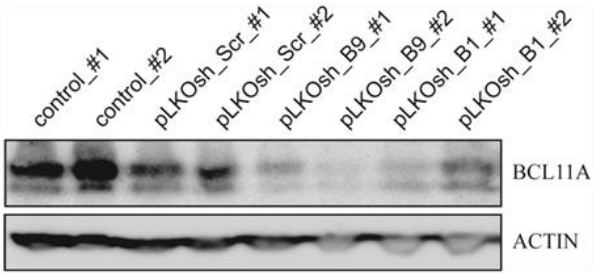
	HBG protein expression			HBB protein expression		
	Intensity (normalized to GAPDH)	fold change (to induced control)	Δ%	Intensity (normalized to GAPDH)	fold change (to induced control)	Δ%
uninduced control_#2	0.00			0.00		
induced control_#2	2.82	1.00		0.86	1.00	
MA821_#2	3.16	1.12	11.99	50.11	57.94	5694.07
MA821Q_#2	4.82	1.71	71.11	13.58	15.71	1470.56
MA821Qsh_G_#2	4.57	1.62	62.36	36.98	42.76	4176.08
MA821Qsh_B1_#1	2.67	0.95	-5.38	0.08	0.09	-91.23
MA821Qsh_B1_#2	1.65	0.58	-41.60	0.20	0.23	-76.57
pLKOsh_Scr_#2	3.82	1.36	35.70	0.39	0.45	-55.09
pLKOsh_B1_#1	1.74	0.62	-38.25	0.26	0.30	-70.38
pLKOsh_B1_#2	2.03	0.72	-28.09	0.09	0.11	-89.37
pLKOsh_B1_(8-1)_#1	2.67	0.95	-5.08	0.06	0.07	-92.56
pLKOsh_B1_(8-1)_#2	2.79	0.99	-0.92	0.13	0.16	-84.47
MA821Q_&_pLKOsh_B1_(8-1)	1.28	0.45	-54.56	0.62	0.72	-27.93

Figure 3.3.14. Expression levels of BCL11A, HBG and HBB protein in HEL cells by Western blot analysis. HEL cells were transduced with LVs to give an estimated VCN of 5 as indicated, and treated with 50 μM haemin. Western immunoblots were probed with antibodies against BCL11A (*Santa Cruz*), HBG, and HBB in whole cell extracts from $4x10^5$ cells. GAPDH was used as internal control. Analysis was performed at day 7 post-transduction. The table shows expression ratio, fold change relative to the control (induced, untransduced) sample, which is arbitrarily set at 1, and the percentage change (Δ%) in expression between the transduced and the control cultures. Molecular weight of BCL11A isoforms XL/L/M/S/XS: 125/100/51/35/25 kDa.

3.3.4 Testing pLKO.1 BCL11A-shRNA LVs in HEK 293T cells

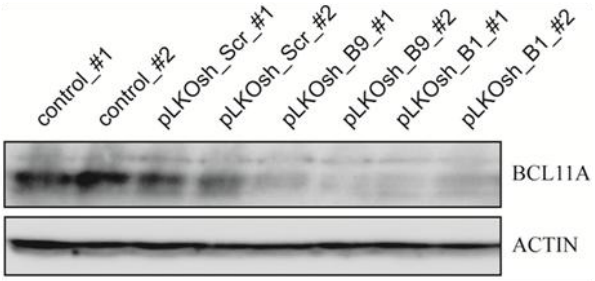
In a parallel study by others in our laboratory, BCL11A protein expression was measured in various available cell lines, namely K562, HEL, MEL, and HEK 293T. Western blot analysis revealed strong signal intensity for bands corresponding to the XL and L isoforms of BCL11A in HEK 293T whole-cell lysates using the anti-BCL11A antibody (*Abcam*). As a result, the HEK 293T cell line was used in a final effort to test the efficiency of U6-driven vectors by measuring BCL11A protein levels. Briefly, HEK 293T cells ($0.5x10^6$) were transduced overnight in duplicate at an estimated MOI of 30. Western blot analysis was performed with whole-cell lysates from $3.5x10^5$ cells at 3-days post-transduction for BCL11A protein expression using the same anti-BCL11A antibodies from both *Abcam*, as suggested by in-house studies, and compared to the BCL11A antibody from *Santa Cruz* in order to be comparable to the work performed with other cell lines. Non-specific knockdown of BCL11A was assessed using control shRNAs. BCL11A knockdown was detected only in cultures transduced with BCL11A-shRNA expression vectors, confirming the potency of the two publically validated BCL11A siRNA target sequences. In comparison to the control, we detected >50% reduction in BCL11A protein levels using either of the *Abcam* (**Figure 3.3.15A**) or *Santa Cruz* (**Figure 3.3.15B**) anti-BCL11A antibodies.

A



	BCL11A protein expression *Intensity (normalized to ACTIN)*		
	ratio	fold change (to control_#1)	Δ%
control_#1	0.82	1.00	
control_#2	1.43	1.76	75.58
pLKOsh_Scr_#1	0.94	1.15	15.35
pLKOsh_Scr_#2	0.88	1.08	8.22
pLKOsh_B9_#1	0.40	0.48	-51.54
pLKOsh_B9_#2	0.23	0.29	-71.45
pLKOsh_B1_#1	0.39	0.47	-52.60
pLKOsh_B1_#2	0.75	0.92	-8.30

B



	BCL11A protein expression *Intensity (normalized to ACTIN)*		
	ratio	fold change (to control_#1)	Δ%
control_#1	0.89	1.00	
control_#2	1.2	1.34	34.12
pLKOsh_Scr_#1	0.9	1.01	0.96
pLKOsh_Scr_#2	0.77	0.86	-14.45
pLKOsh_B9_#1	0.37	0.41	-58.63
pLKOsh_B9_#2	0.17	0.19	-80.71
pLKOsh_B1_#1	0.22	0.25	-75.45
pLKOsh_B1_#2	0.34	0.38	-62.16

Figure 3.3.15. Western blot analysis of BCL11A protein levels in HEK 293T cells. HEK 293T cells were transduced in duplicate at a MOI of 30, as indicated, and analysed for BCL11A protein expression 3 days later. Western blots from whole cell extract ($3.5x10^5$ cells) were analysed with **A)** *Abcam* (1:1000) and **B)** *Santa Cruz* (1:500) antibodies against BCL11A. ACTIN was used as internal control. Values are normalized to the control_#1 sample whose value was arbitrarily set at 1.00. The tables show expression ratio and the percentage change (Δ%) in expression between the transduced and the control cultures. Molecular weight of BCL11A isoforms XL/L/M/S/XS: 125/100/51/35/25 kDa.

3.3.5 Discussion

BCL11A is a transcription factor essential for repressing *HBG* expression and inhibition of which has recently been recognised as a promising target for HbF reactivation. The main goal of this study was to evaluate the knockdown of *BCL11A* expression by RNAi and assess LV function through induction of HBG chain synthesis in mammalian cell-based systems.

The human myelogenous leukemic cell line K562 possesses an embryonic/fetal *HBB*-like globin expression pattern and has been a useful *in vitro* model to study the molecular mechanism of *HBG* expression. Treatment with hydroxyurea (150 μM) has been shown to induce erythroid differentiation [382, 383]. However, a 7-day induction of erythroid differentiation had moderate effects on the K562 cell line stock used in this study, resulting in a low degree of haemoglobinization. The exception was cultures transduced with U6-based pLKO.1 BCL11A_449-shRNA vectors in which a larger fraction of benzidine-positive cells associated with concomitant increase of HBG mRNA and polypeptide chains. In addition, the role of BCL11A in our K562 cell line was not elucidated. Detection of BCL11A protein was ambiguous owing to the presence of persistent background bands on the Western blots. We showed expression of the shorter BCL11A variants of which their individual roles in globin gene expression remains as yet unexplored and found that the larger BCL11A variants, which have been implicated in HBG repression, were absent. Similarly, at the mRNA level, primers amplifying different BCL11A siRNA target sequences shared by all known BCL11A transcripts gave different results. One primer set failed to produce detectable amplicons while the other gave BCL11A mRNA levels comparable to those of controls. Strikingly, the missing sequence in the K562 transcriptome included the target site of the BCL11A_449-shRNA. The absence of the BCL11A region may not be surprising, as the K562 cell line is characterized by neoplastic features and several chromosomal abnormalities [251, 363]. The BCL11A_449-shRNA design is published and shown to work efficiently [63, 119]. However, our study provided no evidence to suggest that BCL11A knockdown played a role in upregulating *HBG* expression.

Moreover, we showed expression of low HBB mRNA and protein levels in naive cultures of K562 cells, which indicated that the *trans*-acting factors necessary for *HBB* expression were

present in the K562 cells used in this study. Transductions of *HBB*-based MA821Q-derived vectors had a robust effect on *HBB* expression, indicating that the RNAi strategy of inserting the shRNA sequence within IVS-II of the *HBB* cassette did not affect transcription and pre-mRNA splicing of the primary transcript. A recent publication [384] demonstrated that transfection of BCL11A-XL or KLF1 plasmids in a K562 clone lacking endogenous expression of these two transcription factors led to moderate levels of HBB transcripts. Further co-tranfection with plasmids encoding BCL11A-XL and KLF1 achieved a prominent increase in HBB mRNA levels, which was annulled when replacing full-length BCL11A-XL with BCL11A-L or BCL11A-S isoforms. Drawing on these findings, we speculated that full-length BCL11A-XL could have been expressed weakly at levels insufficient to repress *HBG* and induce adult *HBB* expression and which fell below the detection threshold of the Western blot assay. The position *HBG1/HBG2* and *HBB* within the *HBB* locus creates competition between the two sets of genes for physical interaction with the upstream βLCR and favours transcription of the more proximal *HBG* (**section 1.4.3** for 'looping model') during the fetal stage of development. We demonstrated that the expression of the *HBB* transgene at low VCN per cell did not affect basal levels of endogenous *HBG* expression. Overall, no difference in HBG mRNA and protein levels was found in K562 cell transduced with the *HBB* transgene-based BCL11A shRNA MA821Qsh_B9 and the control vectors. These findings reflect a non-functional RNAi strategy since these experiments in K562 cells failed to show functionality of the BCL11A-shRNA tools.

Consequently, the human erythroleukemia (HEL) cell line, which has been shown to express the full-length BCL11A-XL, was employed as an alternative cell-based system to assess BCL11A shRNA-expressing vector function. RT-qPCR analysis confirmed that both BCL11A siRNA target sequences were retained in our HEL cells. Similar to that in K562 cells, the HEL cell stock used in this study contained the *trans*-acting factors necessary to regulate *HBB* transgene expression. We showed that none of the BCL11A shRNA-expressing vectors led to a detectable increase in *HBG* expression at levels higher than those of control vectors with concurrent decrease in BCL11A expression. Strikingly, the levels of target gene mRNA and protein did not concur regardless of whether using the same or different reference (*GAPDH* or *ACTIN*) genes for the normalization of target gene expression. Importantly, there was no detectable shRNA

activity from the U6 promoter-based vectors, which are published and shown to work efficiently at least in adult erythroid progenitor cells [119].

RNAi is used to knockdown gene expression and cannot bring about total transcriptional silencing. Therefore the size of the silencing effect on a measured phenotype can be moderate or weak for many potent siRNAs. The failure to detect BCL11A knockdown with concurrent reactivation of *HBG* expression in the K562 and the HEL cells led to the dismissal of these cell lines as a valid platform for vector assessment. Albeit late in the project, the U6 promoter-based vectors were tested in the non-erythroid HEK 392T cell line, which was shown to express the large BCL11A isoforms. Results confirmed the silencing capability of both BCL11A-shRNAs tested here and demonstrated that changes in BCL11A expression can be detected by Western blot analysis.

3.4 Small-volume collection of thalassaemic PBMCs to set-up a biobank

3.4.0 Aims

- To achieve high yield, viability and recovery of leukocytes obtained from small-volume PB samples before and after cryopreservation in 10% DMSO-based media supplemented with 30%, 50%, and 90% FBS.
- To create a repository of leukocyte samples as a source of haematopoietic progenitors.

3.4.1 Introduction

Haematopoietic stem and progenitor cells (HSPCs) are the most extensively studied tissue-specific stem cells with direct application in the clinic for the treatment of blood-related genetic diseases, hereditary immunodeficiencies and haematologic malignancies using haematopoietic stem and progenitor cell transplantation (HSCT) and cell replacement therapy protocols. Paramount to their therapeutic success is their unique engraftment and homing function with ability to repopulate the whole bone marrow (BM) compartment and reconstitute the blood and immune systems after patient chemotherapeutic conditioning. These properties, together with their easier accessibility in high quantities compared to other adult stem cells put HSPCs at the forefront of regenerative medicine [11], ahead also of induced pluripotent stem (iPS) cells and embryonic stem (ES) cells. For purely cell-based therapies, adult HSPCs share with ES and iPS cells the need to identify healthy, compatible donors. However, for clinical applications their lineage restriction and more restricted capacity for indefinite self-renewal, apparent in *in vitro* cultures, and the limited manipulations and culture time required for their use compared to ES and iPS cells are safety features against treatment-related adverse events. Moreover, HSPCs as a substrate for *ex vivo* gene-therapy applications are ethically uncontroversial and readily

available for sufferers of haematopoietic disorders, in contrast to ES cells, and in contrast to iPS cells, HSPCs have a long safety record in clinical application of conventional HSCT, such as BM transplantation [385, 386].

Common sources of HSPCs include extra-embryonic tissues (e.g., umbilical cord blood, CB) and adult tissues (e.g., BM, peripheral blood (PB) and mobilised PB (mPB) after chemical mobilisation of HSPCs form the BM) [387], from which they can be isolated, e.g. by positive selection using monoclonal antibody-bound magnetic immunoaffinity labelled beads directed against the $CD34^+$ cell-surface expression marker, although other markers and their combinations can be used. The $CD34^+$ antigen was reportedly detected in ~1% of the light-density cells isolated from BM and CB, while constituting only <0.01% of unmobilised PB cells [388]. The isolated $CD34^+$ population can be sub-fractioned based on the expression of the transmembrane glycoprotein CD38, with the $CD34^+/CD38^-$ sub-population representing long-term repopulating stem cells and comprising ~5% of the total $CD34^+$ fraction [389]. A recent study of $CD34^+$ subpopulations has revealed their differential suitability for lentiviral gene therapy and highlighted adjustments of transduction protocols for improved manipulation of the $CD34^+/CD38^-$ population [390].

BM has been a classic source of stem cells, while improvements in mobilisation protocols with lowered impact on patients with splenomegaly have brought mPB to the fore [240, 391, 392]. Allogeneic HSCT currently represents the only curative therapy for most inherited blood disorders, while carrying a significant risk of immunological complications (**section 1.3.4.3**) [393]. The alternative of using autologous transplantation of gene-corrected HSPCs, which would minimise immunological complications and remove the need for a compatible donor, has spurred the development of a new form of therapy. As key factors to success, the procurement of a sufficient number of HSPCs (target cell dose: $\geq 6x10^6$ $CD34^+$ cells/kg [394]), efficiency of HSPC gene transfer, homing and engraftment and sufficient levels of conditioning, allowing high donor chimerisms at moderate cell doses, have been identified [240, 395]. In addition to BM and mPB, CB is also considered as a source of HSPCs [396], with CB- and mPB-derived cells being comparable in haemoglobin type and gene expression [397], but with limited cell numbers and availability of CB samples for adult patients [398]. Compared to BM-derived transplants,

collection of mPB HSPCs avoids invasive surgical procedures, speeds up post-procedure recovery and engraftment, reduces the incidence of iatrogenic infections [399] and is thus more attractive to patients. The therapeutic potential of mPB is based on an increase of the usually low $CD34^+$ numbers in PB ($5.6x10^3$ $CD34^+$/mL [400] to upwards of $1.01x10^5$ $CD34^+$/mL (CD34, mobilised) in normal PB stem cell donors) through the application of reagents that induce HSPC mobilization from extravascular haematopoietic sites into the circulation, combined with their collection by leukapheresis [400, 401]. For a long time, granulocyte colony-stimulating factor (G-CSF) was the mobilising agent of choice [240], but its use has been associated with certain morbidities (e.g., hyperleukocytosis with splenic enlargement and thromboembolic events) which prompted re-evaluation and modification of mobilisation protocols, including the use of current protocols and consideration of other or combinations of mobilizing agents to assure high cell yield both cost-effectively and safely [394].

From a practical point of view PB is the largest, cheapest and most accessible of the HSPC sources with several reports documenting the presence of distinct erythroid progenitors in the PB mononuclear cell fraction (PBMC). The constituting HSPCs have the potential to differentiate into mature functional cell types provided that the appropriate culture conditions are applied [402]. This was exemplified by the successful development of *in vitro* culture techniques (tailored erythropoietic growth factor and cytokine composition) which support HSPC expansion, erythroid differentiation and erythroblast production [403]. More specifically, the HSPC-enriched buffy coat fraction, a by-product of standard blood donations, represents an easily accessible and viable stem cell source. The development of extraction techniques which sustain the erythroid potential of the constituting $CD34^+$ cell population would prove highly cost-effective for the establishment of an HSPC reservoir. This was demonstrated in a recent study, which reported the production of a large number of functional RBCs from buffy coats of normal and rare-anaemia blood samples by established *in vitro* culture systems and which demonstrated the preservation of HSPC multi-differentiation potential after cryopreservation of PBMCs [404].

In the present study we investigated the suitability of small-volume pre-transfusion blood samples (2–4mL), such as diagnostic samples received routinely by our department from the Cyprus Thalassemia Centre, to create repositories of leukocyte samples as a source of

haematopoietic progenitors. The establishment of a biobank for these samples, with associated haematological and genotype data, woud provide a comprehensive collection of HSPCs with different genotypes and phenotypes and without additional effort by the collaborating clinicians. Storage of several same-patients samples over time would also allow larger-scale experiment by combination of samples. Here we performed a side-by-side comparison of two cell extraction protocols – i) a density-gradient technique using commercial Ficoll-Paque PLUS or Lymphoprep media and ii) a whole-blood lysis method using self-made ACK lysis buffer, which is comparable to commercially-available buffers [405, 406]. The isolated leukocyte fractions were frozen in 10% DMSO [407, 408] diluted in different concentrations of fetal bovine serum (FBS). The major focus of this study was to establish whether the cryopreserved samples were similar to those that are freshly collected in several characteristics, including yield, viability, haematopoietic potential and transducibility with lentiviral vectors. Reports to date include parallel comparisons of commercially available whole-blood lysis buffers [405, 406, 409-411] or density-gradient media [412-415], as well as comparisons between the two different cell extraction methods [416-420]. The latter studies mainly used flow cytometry immunophenotyping to evaluate the extent to which the process of cell extraction would affect the light-scattering properties and surface marker expression of lymphocyte subsets. To our knowledge, no other study has as yet compared cell extraction methods or cell cryopreservation in terms of the diversity and frequency of haematopoietic progenitors in the isolated leukocytes.

3.4.2 Experimental set-up

In order to evaluate the effects of two density-gradient media (Ficoll-Paque PLUS and Accu-Prep Lymphocyte (Lymphoprep)), one RBC Lysing solution (ACK) and a modified density-gradient protocol (Ficoll-Paque PLUS with additional ACK Lysis wash) on PB processing in combination with freezing and/or lentiviral transduction, the study was designed as depicted in **Figure 3.4.1**:

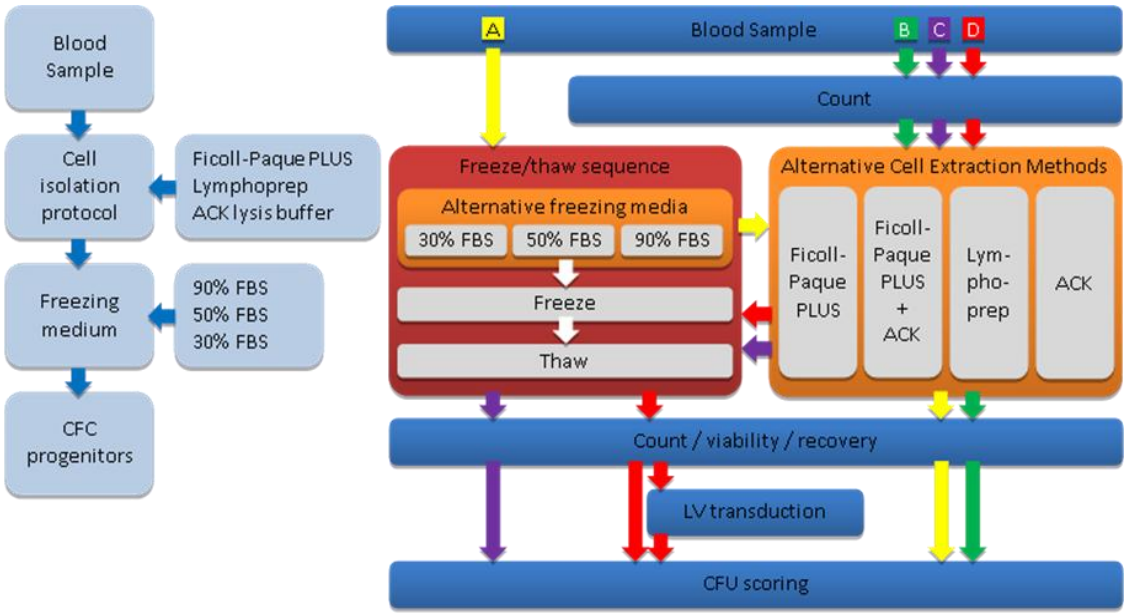


Figure 3.4.1. The schematic of the study consists of four processing paths (A–D), represented by differently coloured arrows. For paths A–C, the whole blood sample was processed using one extraction protocol at a time. The final product was either used for immediate application (fresh, B) or frozen for cryostorage (cryopreserved) before (A) after (C) extraction. For path D, blood samples designated for fresh use were split in half for same-sample assessment of ACK lysis and Lymphoprep extraction. Not all indicated alternative freezing media and cell extraction methods were tested for each pathway (see main text).

3.4.3 The effects of cell isolation and freezing on leukocyte counts

Blood samples were run on a haematology analyser (Sysmex XT-2000i) for white blood cell (WBC) counts and other parameters prior to processing with density-gradient media or an ACK lysis buffer. Isolated mononuclear cells (MNCs) and nucleated cells (NCs), respectively, were counted by trypan blue exclusion on a haemocytometer, before freezing in 1 mL 10% DMSO media supplemented with 30%, 50% or 90% FBS. Similarly, cells were counted after thawing using trypan blue exclusion. Raw data are presented as 'cell counts per mL' instead of 'total cell counts' to account for differences in initial sample volumes (as received from the referring clinicians and usually in the range of 1.5–3 mL).

In order to reduce processing time and cost, blood samples were initially frozen in 10% DMSO with 90% FBS (not shown in **Figure 3.4.1**). To assess sample quality, MNCs were then isolated from thawed blood samples by Ficoll-Paque PLUS density-gradient centrifugation (**Figure 3.4.2A**), revealing an ~87% cell loss in comparison to fresh blood preparations. Cell counting on the haemocytometer presented a challenge owing to the large number of RBCs and platelets in the isolated MNC fraction as a result of poor phase separation by centrifugation. To address this while also improving precision and reproducibility in counts, the MNC fraction was treated with an ACK lysis wash (**Figure 3.4.2A**). Owing to substantial cell loss by freezing of whole blood and to substantially increased cryostorage-space requirements compared to other methods, this freezing protocol was not pursued further. Instead, fresh blood samples were processed with Ficoll-Paque PLUS density-gradient, and MNCs were frozen in 90% FBS-based freezing media (**Figure 3.4.2B**), resulting in an acceptable ~32% decrease in average MNC counts compared to fresh blood.

Recently, we adopted an *in vitro* two-phase erythroid liquid culture system which uses Axis-Shield's Lymphoprep^{TM} density-gradient (1.077±0.001 g/mL) for the isolation of MNCs from fresh PB samples. In order to standardise protocols and use the same density-gradient medium for both experimental protocols, we investigated the suitability of Lymphoprep to replace Ficoll-Paque PLUS. First, the effect of the two density-gradient media was compared on fresh MNCs. Lymphoprep (14.82±7.42x10^5 MNC/mL, n=5) resulted in 65% lower MNC counts compared to

the Ficoll-Paque PLUS (42.34±37.7x10^5 MNC/mL, n=2), which owing to sample size and variation was non-significant (t=1.814, *p*=0.1293) (**Figure 3.4.2C**) and is not in line with published results [412]. Second, the effect of freezing on Lymphoprep-isolated MNCs was investigated using a 30% and a 50% FBS-based freezing medium. Freezing of Lymphoprep-isolated MNCs in 50% FBS was the standard procedure for $CD34^+$ cryostorage according to the liquid culture protocol (Protocol C). Freezing in 30% FBS was investigated on the basis that if shown to be equally efficient to a 50% FBS freezing, it would prove cheaper long-term for cryostoring numerous small samples. Analysis revealed a non-significant freezing effect on average MNC counts compared to fresh samples (F(2,12)=0.2316, *p*=0.8689). In fact, freezing in 30% FBS resulted in 0.3% lower counts whereas freezing in 50% FBS resulted in 12% lower counts (**Figure 3.4.2D**). Moreover, freezing of Ficoll-Paque PLUS-isolated MNCs in 90% FBS (28.77±30.01x10^5 MNC/mL, n=6) did not produce significantly different MNC counts from freezing of Lymphoprep-isolated MNCs in 50% (13.03±5.72x10^5 MNC/mL, n=6) or 30% FBS (14.77±5.34x10^5 MNC/mL, n=4), (F(2,13)=1.170, *p*=0.341) **Figure 3.4.2D**. Overall, these data showed that, in terms of MNC numbers, either density gradient media would be equally effective to use on fresh preparations. In addition, freezing of the MNC fractions did not affect the number of viable MNCs separated using Lymphoprep or Ficoll-Paque PLUS.

The Lymphoprep cell isolation protocol presented a good record of reproducibility and precision. Nonetheless, it is time-consuming and costly when dealing with many small-volume blood samples. Therefore we considered the alternative of a self-made ACK buffer for RBC lysis and freezing of the resulting NCs, a procedure which is inexpensive, simple and quick. First, we addressed the question whether a simple RBC lysis would be equally efficient to a density-gradient technique. Compared to the Lymphoprep density-gradient medium (14.82±7.42x10^5, n=5), the ACK lysing buffer (24.10±12.7x10^5, n=8) resulted in ~63% higher cell numbers per mL on fresh preparations, which was nonetheless not considered significant (t=1.472, *p*=0.1690). The observed trend was expected, as the ACK-isolated NC fraction comprises MNC and polymorphonuclear (PNC) leukocytes, while the Lymphoprep density-gradient centrifugation physically separates both cell types and allows the sampling of enriched MNCs [421]. Assuming an approximate ratio of 2:5 for MNCs to total nucleated blood cells (for [422]), ACK Lysing

performed comparably to the Lymphoprep density-gradient technique in isolating leukocyte fractions from fresh blood samples. Secondly, the effect of freezing on ACK-isolated leukocytes was investigated using a 30%, 50% and 90% FBS-based freezing medium, in line with media used to freeze MNCs isolated by density-gradient protocols. We detected a significant difference in average NC counts per mL using only the 50% FBS freezing medium ($F(3,16)=4.553$, $p=0.0173$) (**Figure 3.4.2E**), once more likely because the analysis was underpowered owing to sample-to-sample variation and small sample numbers. However, we saw the expected trend of lower cell recovery rates compared to fresh sample for lower FBS concentration in the freezing medium, specifically down to 64.2% with 30% FBS ($8.6\pm6.99x10^5$; n=4), to 65.9% with 50% FBS ($8.2\pm4.9x10^5$; n=6) and to 5% with a 90% FBS ($22.9\pm4.1x10^5$; n=2), with a difference of >50% between 30% and 90% FBS. Next, we performed a side-by-side comparison of mean cell counts between fresh and frozen preparations using either ACK Lysing or Lymphoprep density-gradient separation. With exception to a significant loss in NC numbers after a 50% FBS freezing, we did not detect significant differences across any of the other mean scores ($F(5,27)=3.259$, $p=0.0198$), suggesting that any observed variation between the sample means was likely due to chance and that additional samples might allow a clearer interpretation of results (**Figure 3.4.2F**). Similarly, analysis showed that there was no significant difference in the number of viable leukocytes in frozen samples (90% FBS-based freezing) using Ficoll-Paque ($28.77\pm30.0x10^5$, n=6) and ACK ($22.90\pm4.10x10^5$, n=2), $t=0.288$, $p=0.782$, in the absence of further data once more suggesting that ACK-based lysis and density-gradient isolation would result in equivalent cell numbers. However, absence of proof is not proof of absence, and even with the limited data in hand and pending further analyses, a trend of improved recovery with higher FBS concentration in the freezing medium and of a decreased viability for frozen cells after ACK-based lysis is apparent.

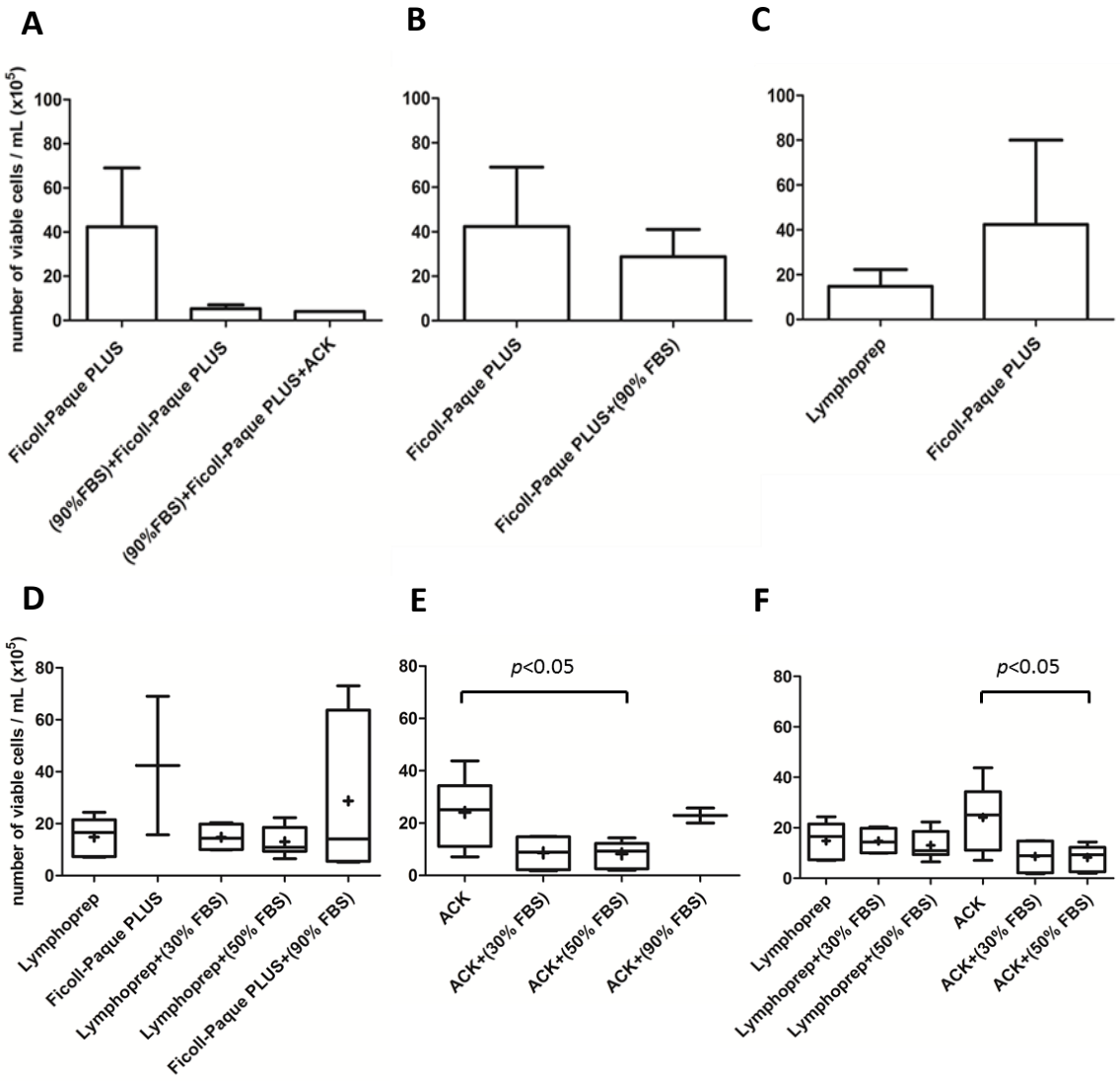


Figure 3.4.2. Blood processing conditions (cell extraction and/or freezing) and their effect on leukocyte counts. A–C) Histograms representing means ± SD. **D–F)** Box and whisker plots showing median (line across boxes), mean ('+' in the middle of the boxes), 25% to 75% interquartile range (boxes), and maximum and minimum values (the end of the vertical line across the boxes). Statistical analysis was performed by t-test (n=2) and one-way ANOVA (n>2). Statistically different comparisons are annotated (horizontal brackets, $p<0.05$).

3.4.4 The effect of cell isolation and freezing on cell recovery and viability

In addition to its effect on raw leukocyte counts, we next investigated the effect of blood processing on cell viability and recovery after extraction (fresh preparations) and after thawing (frozen preparations). Recovery scores (yield of viable cells) were calculated by dividing the total number of live cells in the isolated cell fraction by the total number of cells in the harvested blood and multiplying by 100 to obtain percentage recovery. Similarly, viability scores were calculated by dividing the number of live cells per mL at the end of the blood processing by the number of all cells (alive and dead) per mL and multiplying by 100 to obtain percentage viability. Comparing the effect of ACK Lysing (58.28±39.70, n=6) and Lymphoprep density-gradient (45.58±29.49, n=6) separation on fresh preparations did not show a significant difference in mean recovery scores (t=0.6286, *p*=0.5437). However, Lymphoprep-isolated leukocytes (88.56±7.76, n=6) exhibited a significantly higher viability by ~18% compared to ACK-isolated leukocytes (75.27±8.94, n=6) (t=2.752, *p*=0.0204).

Extending the analysis to frozen preparations, we did not detect a significant difference in the percentage recovery or the percentage viability of Lymphoprep-isolated leukocytes (**Figure 3.4.3A and 3B**). Compared to the fresh MNC fraction (45.58±29.49, n=6), cryostorage with either 30% FBS freezing medium (31.93±24.71, n=4) or a 50% FBS freezing medium (17.93±7.62, n=6), resulted in a 70% and 39.3% recovery respectively (F(2,13)=2.306, *p*=0.1390). Similarly, cryostorage with either a 30% FBS freezing medium (78.19±14.06, n=4) or a 50% FBS freezing medium (74.21±11.24, n=5) resulted in an 88.3% and 83.8% viability compared to the fresh MNC fraction (88.56±7.76, n=6) (F(2,12)=2.596, *p*=0.1156). Of note, all blood samples were taken from people with thalassaemia just prior to transfusion, collected and processed on different days. These samples were shown to vary in blood volume and leukocyte counts, and were randomly allocated for processing. Furthermore, the density-gradient separation is a labour-intensive process requiring skilful handling, and as such could result in operator variability with regard to buffy coat fraction collection and thus cell recovery. Despite a visible trend from the data suggesting lower percentage recovery for cryostored MNC samples, the level of variability within groups once more prevented a statistically significant result.

Moreover, analysis showed that freezing of ACK-isolated NCs would influence the percentage recovery ($F(2,13)=4.992$, $p=0.0246$) (**Figure 3.4.3C**). Post-hoc Tukey's test showed the significance (α-level=0.05) of a 79.6% lower recovery for cells cryopreserved in 50% FBS-based freezing medium (11.89±10.02, n=6) compared to fresh cells (58.28±39.70, n=6). A 30% FBS-based freezing (17.69±20.04, n=4) resulted in a 69.6% lower recovery which was not considered significant. Further analysis showed that freezing of ACK-isolated NCs did not affect percentage viability significantly ($F(2,12)=2.102$, $p=0.1686$) (**Figure 3.4.3D**). Compared to fresh NCs (75.27±8.94, n=6), cryostorage with a 30% FBS freezing medium (59.0±15.13, n=3) or a 50% FBS freezing medium (72.77±12.25, n=5) resulted in an 78% and 96.7% viability, respectively. These data strongly suggested that a 30% FBS-based freezing medium is better suited for NC cryostorage.

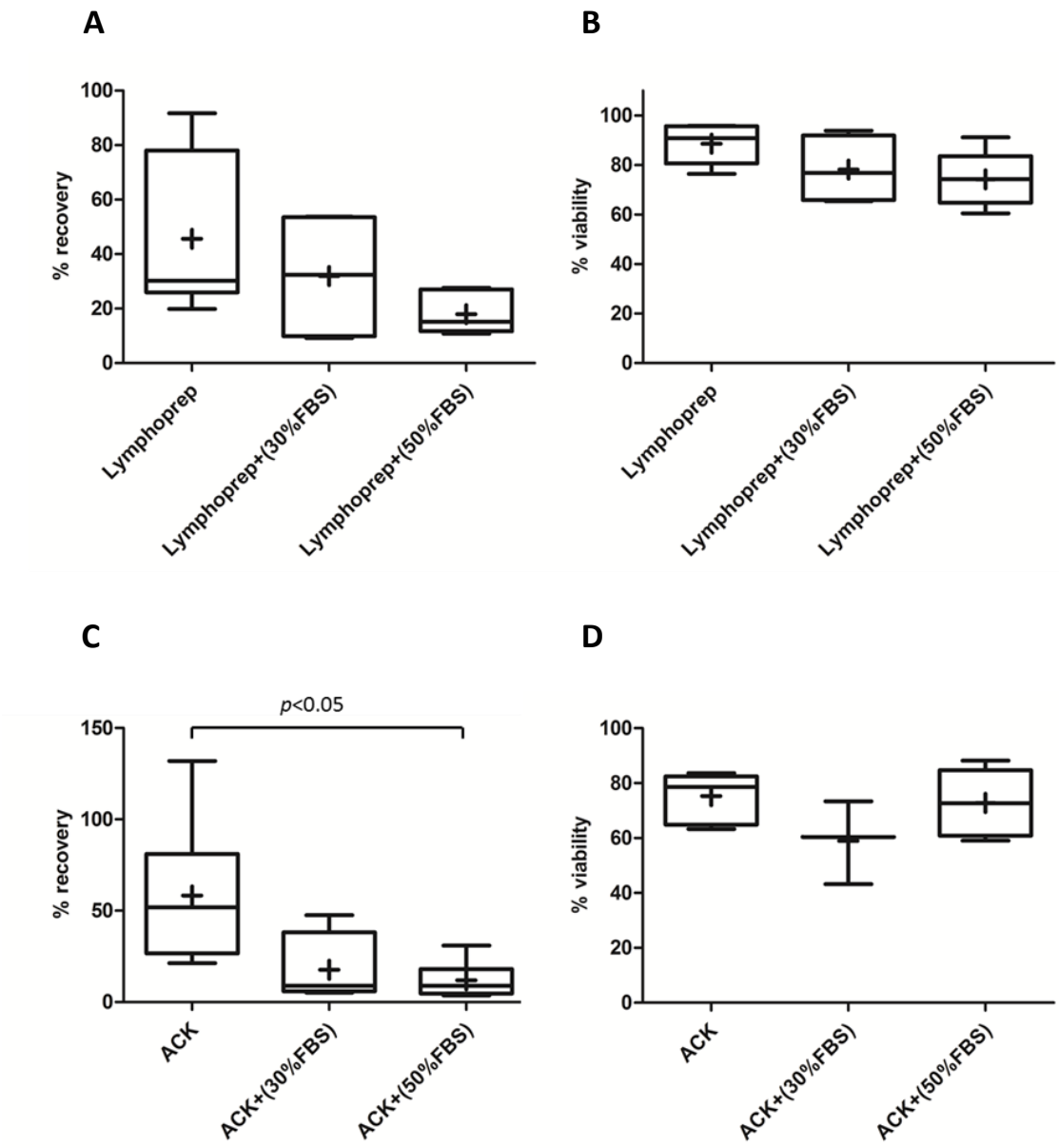


Figure 3.4.3. Blood processing conditions (cell extraction and/or freezing) and their effect on leukocyte recovery and viability. Mean recovery and viability of cells isolated from thalassaemic blood donors using Lymphoprep density-gradient centrifugation (**A–B**) and ACK RBC lysing buffer (**C–D**). Results are displayed in box and whiskers plots, showing median (the line across boxes), mean ('+' in the middle of the boxes), 25% to 75% interquartile range (the boxes), and maximum and minimum values (the end of the vertical line across the boxes). Statistical analysis was performed by one-way ANOVA. Statistically different comparisons are annotated (horizontal brackets, $p<0.05$).

3.4.5 The effects of cell isolation and freezing on colony-forming cell (CFC) frequencies

The effect of different isolation protocols and freezing conditions on progenitor frequencies was assessed using preparations of methylcellulose culture medium (MCM) assays, followed by colony scoring and quantification based on CFU (colony-forming unit) size, morphology, and cellular composition. We classified colonies under three categories, namely BFU-E, CFU-G/M/GM (CFU-M, CFU-G and their common precursor CFU-GM) and CFU-GEMM. The number of colonies provides a measure of the number of viable and functional CFCs (colony-forming cells) in the samples. The CFU and CFC terms have been used interchangeably in the literature to describe *in vitro* functional assays for enumerating and/or characterizing haematopoietic progenitors in semi-solid media in response to cytokine stimulation. In order to avoid any confusion in terminology, in this chapter, 'CFC' refers to PB-derived progenitor cells, which are capable to develop into colonies, while 'CFU' refers to individual colonies of any type. CFC progenitor frequencies (per mL) were calculated by dividing the average colony number by the blood-volume equivalent of the number of live cells that were initially plated, i.e. (number of plated live cells)*(live cells/mL of original blood volume). Of note, we observed a large difference in the raw data for frozen MNCs separated by Ficoll-Paque PLUS which prompted us to apply outlier removal to this data set and also across the other data sets for consistency (**section 2.14**).

3.4.5.1 Effect of the cell isolation method on CFC frequency

While freezing full blood reduced initial processing time and cost, *in vitro* CFU yield from such samples by Ficoll-Paque PLUS density-gradient isolation was below the detection threshold of our experiments (0.00±0.00, n=2). Processing of frozen blood exhibited poor phase separation during centrifugation which resulted in RBC and platelet contamination of the separated buffy coat. In order to clear the contamination, the buffy coat was treated with an ACK wash, which, however, was inefficient, resulting in minimal CFC progenitor yield (2.67±0.03, n=2) and in extremely low CFU numbers (3 BFU-Es and 1 CFU-G/M/GM, n=2). Next, we compared the effect of the Ficoll-Paque PLUS density-gradient protocol and its modified ACK version on the

CFC frequencies of fresh MNCs (**Figure 3.4.4A**). We showed that the unmodified Ficoll-Paque PLUS protocol (727±45.2, n=2) resulted in 92% higher CFC frequencies compared to the modified ACK version (61.5±42.1, n=2), which was tested highly significant even for the small sample investigated (t=15.22, *p*=0.0043). These data showed that the ACK wash of the MNC fraction was detrimental to CFC survival, and thus unsuitable for MNC processing and biobank cryostorage.

In **section 3.4.4**, we proposed the replacement of the Ficoll-Paque PLUS with the Lymphoprep density-gradient medium. We detected a significant difference of 67% in CFC frequencies of fresh MNCs separated by density-gradient using Ficoll-Paque PLUS (727±45.2, n=2) compared to Lymphoprep (240.6±138.5, n=5), (t=4.631, *p*=0.0057) (**Figure 3.4.4A**). The post-thawing frequencies of MNCs separated by Ficoll-Paque PLUS (55.5±56.6, n=5) were reduced to 7.6% of fresh preparations, which was statistically significant (t=14.73, *p*<0.0001) (**Figure 3.4.4B**). On the other hand, the cryopreservation of MNCs separated using Lymphoprep did not influence the frequencies of CFCs consistently, with 30% FBS giving a 50.7% decrease (118.4±87.8, n=3) and with 50% FBS giving a 20% increase (300.8±189.7, n=5) in CFC frequencies, respectively (**Figure 3.4.4C**). The 39 samples for these analyses were analysed and processed independently on different days, so that for the reagents in hand a general recommendation against freezing in 90% FBS after Ficoll-Paque PLUS density-gradient centrifugation and for freezing in 50% FBS after Lymphoprep density-gradient centrifugation can be given. Confirming these findings with independent batches of Ficoll-Paque PLUS and Lymphoprep will help generalise this statement.

3.4.5.2 Effect of the freezing protocol on CFC frequency

We then investigated the effect of freezing in different freezing media on the CFC frequencies of ACK-isolated leukocytes. Analysis showed an overall significant difference in CFC frequencies between fresh and frozen NCs (F(3,11)=5.972, *p*=0.0087) (**Figure 3.4.4D**). A post-hoc Tukey's test showed that the 50% FBS-based freezing media resulted in significantly lower CFC frequencies compared to fresh NCs (532.7±390, n=7). The 30% FBS (20.84±12, n=3) and the 50% FBS (24.5±25.9, n=5) freezing media lowered CFC frequencies to 3.9% and 4.6% of

control levels, respectively. However, the 90% FBS-based freezing medium (677.4±100.1, n=2) did not show significantly different CFC frequencies compared to fresh NCs. These data showed that freezing of NCs using either a 30% or a 50% FBS-based freezing medium had detrimental effects on CFC frequencies, whereas the 90% FBS-based freezing medium was suitable for the cryostorage of ACK-isolated cells.

Lastly, we performed a side-by-side comparison of the CFC frequencies between fresh and frozen leukocyte samples isolated using the density-gradient media or whole-blood lysis methods ($F(5,17)=15.51$, $p<0.0001$). For the reagents used, our analyses recommended freezing in 50% FBS for Lymphoprep-isolated MNCs (300.8±189.7, n=5) and in 90% FBS for ACK-isolated NCs (677.4±100.1, n=2) (**Figure 3.4.4E**).

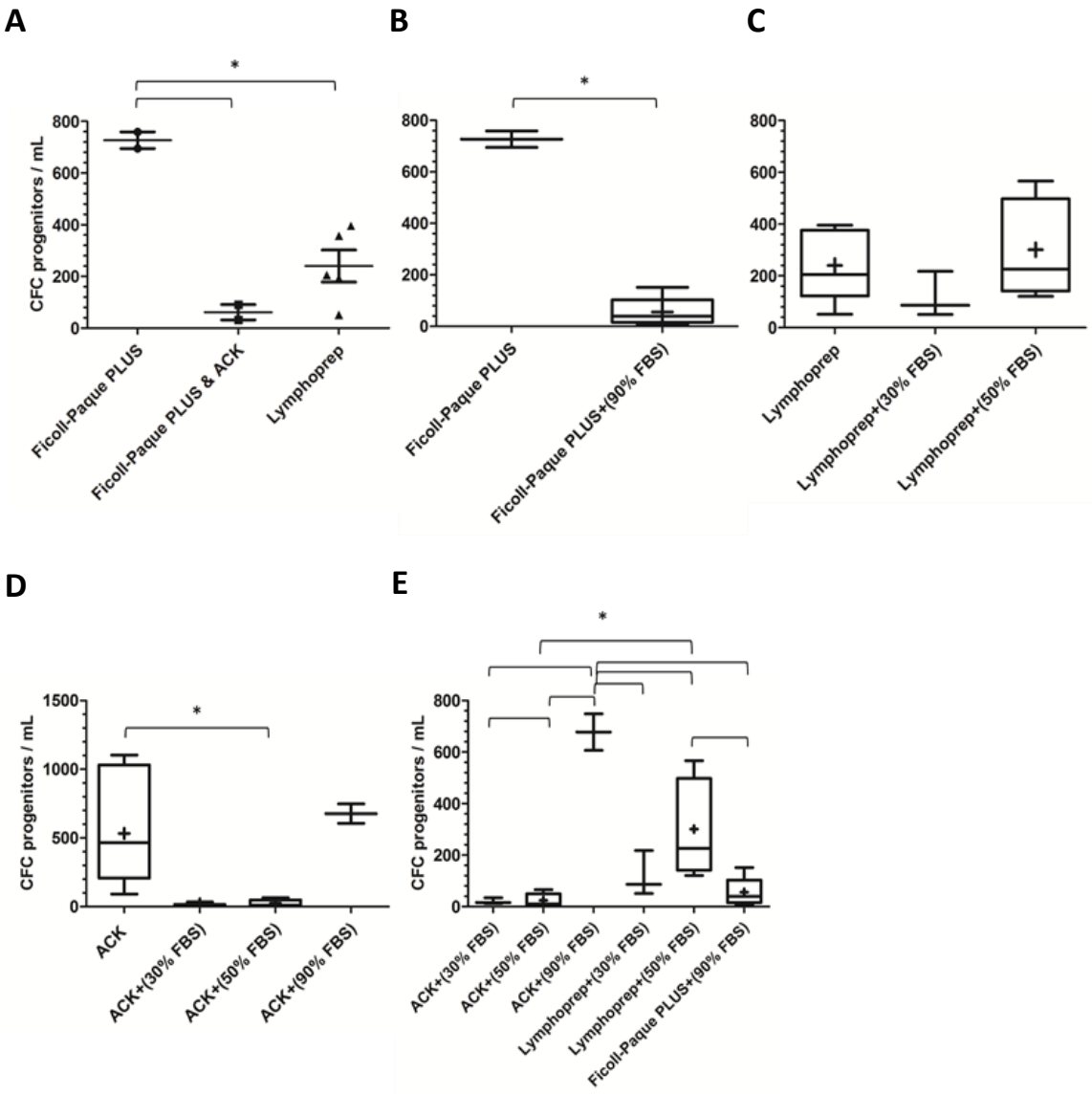


Figure 3.4.4. Blood processing conditions (cell extraction and/or freezing) and their effect on CFC frequencies. Values represent counts of CFC progenitors per mL. **A)** Scatter dot plots show mean ± SD. **B–E)** Box and whiskers plots showing median (the line across boxes), mean ('+' in the middle of the boxes), 25% to 75% interquartile range (the boxes), and maximum and minimum values (the end of the vertical line across the boxes).Statistical analysis was performed by t-test (n=2) or one-way ANOVA (n>2). Statistically different comparisons are annotated (horizontal brackets & *, $p<0.05$).

3.4.6 The effects of cell isolation and freezing on CFC differentiation

Work presented thus far has established the unsuitability of frozen whole blood as a source for high CFC progenitor yield and suggested the extraction of leukocytes from fresh blood and their immediate cryostorage as an alternative strategy for developing a tissue biobank system. The effect of different isolation protocols and freezing conditions on the differentiation potential of extracted CFC progenitors was assessed in CFU assays for enumerating multipotential (e.g., CFU-GEMM) and lineage-committed (e.g., BFU-E and CFU-G/M/GM) haematopoietic progenitor cells. All statistical analyses were performed with a two-way ANOVA test.

Initial attempts to identify an appropriate blood processing method used Ficoll-Paque PLUS density-gradient and a modified version of the protocol which included an ACK wash of MNCs. A two-way ANOVA followed by Bonferroni's post-test showed that these two protocols did not affect the differentiation potential of extracted CFC progenitors in fresh samples ($F(1,6)=0.0$, $p=0.9998$). The Ficoll-Paque PLUS protocol gave 63.1±0.8 BFU-E, 33.8±3.6 CFU-G/M/GM and 3.1±4.4 CFU-GEMM, and its modified version gave 61.5±19.4 BFU-E, 29.4±6.6 CFU-G/M/GM and 9.1±12.8 CFU-GEMM. Instead, we detected a significant difference in the percentage of CFU sub-types, indicating that the differentiation fate of isolated CFC progenitors, accounting for ~90% of the total variance between CFU sub-type means, was critical for the type and number of CFUs scored ($F(2,6)=30.67$, $p=0.007$) (**Figure 3.4.5A**). In addition, the cryostorage of Ficoll-Paque PLUS-isolated MNCs in 90% FBS-based freezing media did not give different number of CFU sub-types (67.9±40.8 BFU-E, 9.2±14.4 CFU-G/M/GM and 6.2±13.4 CFU-GEMM) compared to fresh samples ($F(1,18)=0.24$, $p=0.6282$). Similar to fresh preparations, analysis revealed significant difference in the percentage of CFU sub-types, inherent to the differentiation fate of isolated CFC progenitors ($F(2,18)=10.35$, $p=0.0010$) (**Figure 3.4.5B**). The data showed that frozen and fresh MNCs separated using Ficoll-Paque PLUS maintained similar percentages of CFU sub-types.

We next compared the effect of Ficoll-Paque PLUS and Lymphoprep cell isolation protocols on the percentage of CFU sub-types in frozen MNCs. In comparison to their fresh preparations, neither of the two protocols had a significant effect on the percentage of CFU sub-types in

frozen samples (F(4,60)=0.41, p=0.7978) (**Figure 3.4.5C**). Instead, there was a significant difference in the number of CFU sub-types (F(2,60)=139.25, p<0.0001) with the majority quantified as BFU-Es, for both protocols with either fresh or frozen preparations.

For ACK-isolated NCs, analysis showed a significant interaction effect between NC freezing and differentiation fate of the constituting CFC progenitors (F(6,42)=4.28, p=0.0019) (**Figure 3.4.5D**). A multiple comparison Bonferroni post-test indicated that a 50% FBS-based freezing resulted in significantly different scores for the BFU-E and the CFU-G/M/GM clusters. Compared to the fresh NCs (78.8±10.8 BFU-E, 21.1±11.1 CFU-G/M/GM), the 50% FBS-based freezing (44.4±35.3 BFU-E, 54.7±36.4 CFU-G/M/GM) resulted in a 43.3% lower percentage of BFU-E and a 159.9% higher percentage of CFU-G/M/GM, p<0.05. Compared to the 30% FBS-based freezing (77.2±21.6 BFU-E, 22.8±21.6 CFU-G/M/GM), the 50% FBS-based freezing resulted in a 42.4% lower percentage of BFU-E and a 140% higher percentage of CFU-G/M/GM, p<0.05. Compared to the 90% FBS-based freezing (86.1±11.6 BFU-E, 13.9±11.6 CFU-G/M/GM), the 50% FBS-based freezing resulted in a 48.4% lower percentage of BFU-E and a 293.5% higher percentage of CFU-G/M/GM, p<0.05. This result suggests that a 50% FBS-based freezing favoured the survival and/or proliferation of GMP (granulocyte/macrophage) oligopotent progenitors over erythroid progenitors.

A two-way ANOVA analysis was then performed between fresh and frozen leukocytes separated by ACK Lysis and Lymphoprep density-gradient. Analysis showed a significant interactive effect between the cell extraction protocol and/or leukocyte freezing and the differentiation fate of the constituting CFC progenitors (F(10,78)=4.12, p<0.0001) (**Figure 3.4.5E**). A multiple comparison Bonferroni post-test indicated that the 50% FBS-based freezing of ACK-isolated NCs (44.4±35.3 BFU-E, 54.7±36.4 CFU-G/M/GM) resulted in significantly different percentages of BFU-E and CFU-G/M/GM colonies compared to the other frozen NC samples (**Figure 3.4.5D**), as well as in the frozen MNC samples. More specifically, the 50% FBS-based freezing of NCs resulted in a 44.5% lower percentage of BFU-E and a 175% higher percentage of CFU-G/M/GM compared to the 30% FBS-based freezing of MNCs (80.2±11.2 BFU-E, 19.8±11.2 CFU-G/M/GM), and in a 44.1% lower percentage of BFU-E and a 167%

higher percentage of CFU-G/M/GM compared to the 50% FBS-based freezing of MNCs (79.4±14.6 BFU-E, 20.5±14.6 CFU-G/M/GM), $p<0.05$.

Overall, these findings showed that the cell isolation protocol and/or freezing medium did not bias CFC extraction towards particular CFU sub-types and/or affect the differentiation potential of extracted CFC progenitors in MNC and NC samples. 50% FBS-based freezing of ACK-isolated NCs was an exception and resulted in significantly lower BFU-E scores and as the only ACK-based treatment produced CFU-GEMM (0.8±1.7), in spite of low frequencies (0.83%). The multi-lineage progenitor CFU-GEMM and the erythroid-committed BFU-E are of particular interest in the therapeutic assessment of GT vectors carrying eythroid promoters.

Statistics recommend any combination of Lymphoprep and freezing regimen, while for ACK lysis favouring a 30%-FBS freezing medium for WBC cryostorage with high CFC yield.

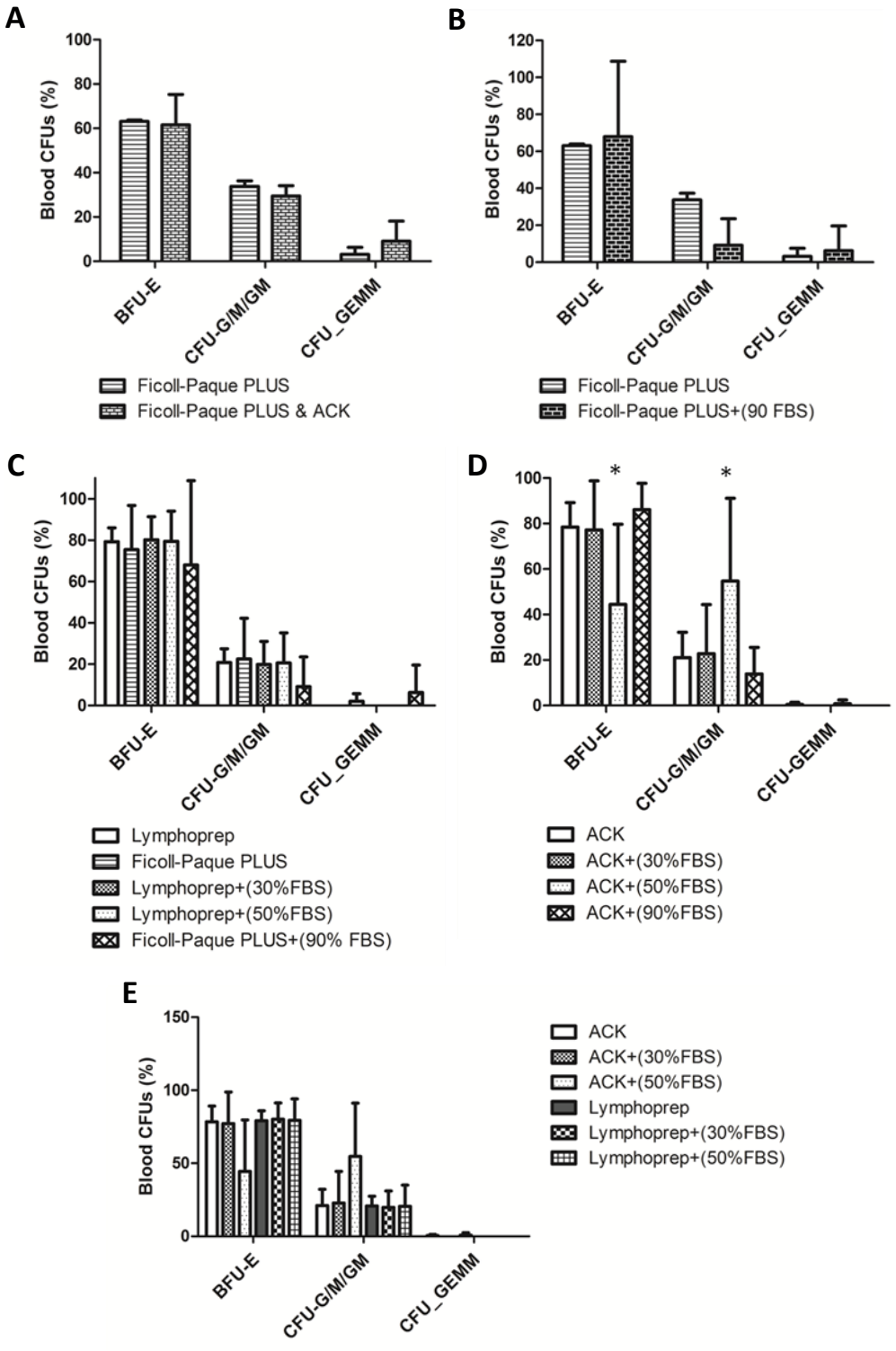


Figure 3.4.5 Blood processing conditions (cell extraction and/or freezing) and their effect on CFU frequencies. A–E) Values represent mean percentage (%) of CFU colonies BFU-E, CFU-G/M/GM and CFU-GEMM for various processing conditions. Histograms represent means ± SD. Statistical analysis was performed by one-way ANOVA. Statistically different values are annotated (*, $p<0.05$).

3.4.7 The effects of blood processing and freezing on CFC transduction efficiency

The effect of viral transduction on CFC progenitor frequencies and its propensity for transduction of particular CFC types was assessed after cell extraction (fresh preparations) and after thawing (frozen preparations). We used GFP reporter gene expression to assess the efficiency of LV gene transfer in CFC progenitors via quantification of green fluorescent CFUs using fluorescence microscopy. The investigation was performed on leukocyte samples from two patients of whom small-volume blood samples were received recurrently within a 6 month period. For each patient, a total of three blood samples were required for experimental set-up. These samples were received on different days and processed using Lymphoprep density-gradient or ACK lysis. Leukocytes designated for cryopreservation were suspended in 50% FBS-based freezing media. Leukocytes from fresh or frozen preparations were cultured in the expansion medium of a pre-established liquid culture system at a high cell density ($1x10^6$/mL) for 24–48 hours in order to enable cell recovery. Subsequently, leukocytes were transduced with a PGK-GFP-expressing vector using standard procedures (**section 2.8.4**) in which $2x10^5$ cells were mixed with vector-containing medium to achieve an estimated VCN of 7.5. As transduced cells were not enriched for $CD34^+$ progenitors, other types of cells (such as non-myeloid progenitors), also susceptible to viral transduction, would have equal chance of being transduced. Therefore we included an additional transduction using 6X the initial VCN; that is, to give an estimated VCN of 45.

The experimental design included four distinct blood processing conditions: i) Lymphoprep, ii) ACK, iii) Lymphoprep+(50% FBS), iv) ACK+(50% FBS). Therefore, the blood (three vials) of one patient was separated into four groups. Each of these was divided into three sub-groups according to transduction: i) mock = untransduced ii) PGK-GFP 1X = transduced with estimated VCN 7.5, iii) PGK-GFP 6X = transduced with estimated VCN 45. Transduction was performed as described (**section 2.9.5.3.3**).

The study was designed as shown in **Figure 3.4.6**.

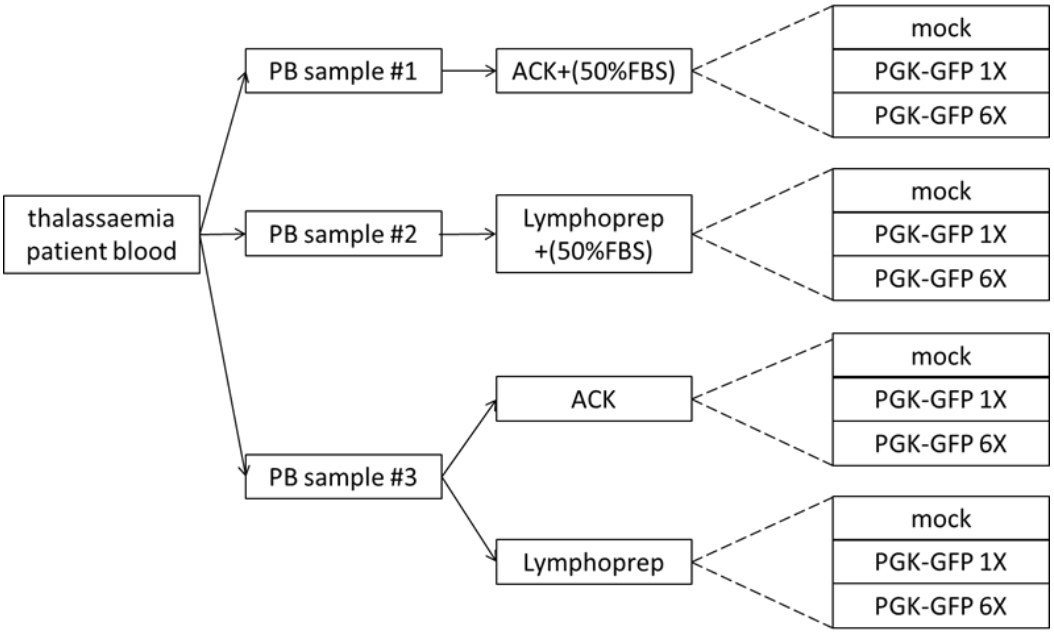


Figure 3.4.6. Diagrammatic illustration of the study. Three same-patient blood samples were received and processed on different days using different protocols. One sample was processed with ACK lysis extraction, followed by freezing in 50% FBS-based medium. One sample was processed with Lymphoprep extraction, followed by freezing in 50% FBS-based medium. One sample was split in half for same-sample assessment of ACK lysis and Lymphoprep extraction. Each of these preparations was sub-divided into three samples of equal cell number for mock treatment and transduction with PGK-GFP LV at 1X and 6X viral inoculum.

Transduced and non-transduced (mock) cells were mixed with semi-solid methylcellulose media and plated into 35-mm dishes, one dish per treatment. CFC progenitor frequencies were calculated from the total CFU count as previously described (**section 3.4.5**). A two-way ANOVA was performed to assess the effect of blood processing (cell extraction and/or freezing) and LV transduction on CFC frequencies. Analysis indicated a non-significant interaction suggesting that the effect of transduction (vs. mock) on CFC frequencies was not different between fresh or frozen samples separated using Lymphoprep or ACK ($F(6,12)=0.42$, $p=0.8523$) (**Figure 3.4.7A**). Further analysis without an interaction term showed that freezing of NCs with 50% FBS-based freezing medium (56.8±31.7, n=2) led to a significant reduction by 83.3% in CFC frequencies compared to fresh preparations (341.1±187.4, n=2) ($F(6,12)=4.46$, $p=0.0356$). Despite a non-statistically significant outcome, the observed difference in CFC frequencies between the transduced and non-transduced samples was large (e.g., we detected a 68.8% and 83.7%

reduction in progenitor frequencies with PGK-GFP 1X and 6X, respectively, compared to mock for ACK+(50% FBS) preparations) (**Figure 3.4.7A**). Sample-to-sample variation and the small sample size per treatment (n=2) led to the increased variance of the sample mean that could result in a wider confidence interval and thus greater uncertainty about the true effect of the investigated parameter. In this respect, instead of using absolute values with high risk of error in statistical hypothesis testing, we calculated the fold change in CFC frequencies ('test' value divided by 'mock' value; non-transduced samples were arbitrarily assigned a score of 1). The observed fold changes are the average of the two independent experiments (**Figure 3.4.7B**). Analyses indicated a significant interaction suggesting that the effect of transduction on CFC frequencies was different between fresh or frozen samples separated using Lymphoprep or ACK ($F(6,12)=4.62$, $p=0.0118$). A Bonferroni's post-hoc test for multiple comparisons identified a significant difference in CFC frequencies for the PGK-GFP 6X treatment of ACK+(50%FBS) compared to i) Lymphoprep ($p<0.001$), ii) ACK ($p<0.01$), and iii) Lymphoprep+(50%FBS) ($p<0.05$). In all three pair-wise comparisons, the ACK+(50% FBS) condition exhibited reduced CFC frequencies by i) 95.3%, ii) 95.5% and iii) 86.5%, respectively. It would appear that the ACK+(50% FBS) samples were least suitable for transductions with large amounts of LV. In addition, we showed that CFC frequencies were similarly affected by high or low viral vector quantities.

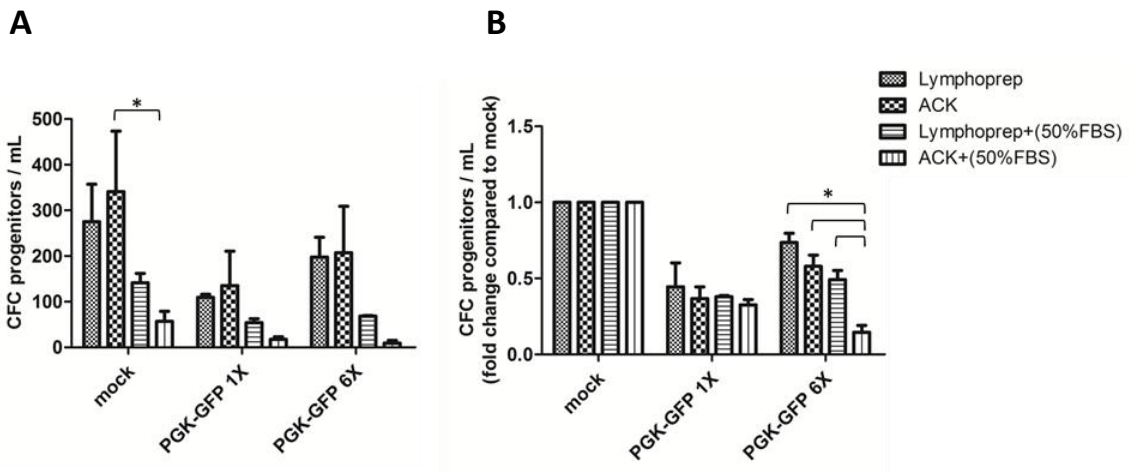


Figure 3.4.7. The effect of LV transduction on CFC frequencies for different processing conditions (cell extraction and/or freezing). Data represent (**A**) absolute values and (**B**) relative fold ratios of CFC frequencies. Histograms represent means ± SD. Statistical analysis was performed by two-way ANOVA. Statistically different values are annotated (horizontal brackets & *, $p<0.05$).

Next, we investigated whether the effect of transduction on CFC frequencies was directed towards a particular progenitor type. Following CFU colony scoring and quantification, we detected a difference in the total number of colonies scored per plate between samples. To allow comparison between samples and instead of the variable absolute value, we therefore used the percentage of each CFU type of the total CFU count. Analysis was performed by a two-way ANOVA. We showed that for ACK-isolated leukocytes, the transduction of fresh or frozen preparations did not influence the percentage of different CFU sub-types ($F(5,18)=0.0$, $p=1.00$). Instead, it was the differentiation fate of CFC progenitors constituting the NCs that significantly affected the percentage of CFUs, accounting for 97.13% of the total variance between sample means ($F(2,18)=397.98$, $p<0.001$) (**Figure 3.4.8A**). Similarly, we showed that for Lymphoprep-isolated leukocytes, the transduction of fresh or frozen preparations did not influence the percentage of different CFU sub-types ($F(5,18)=0.00$, $p=1.000$). Instead, it was the differentiation fate of CFC progenitors constituting the MNCs that significantly affected the percentage of CFUs, accounting for 99.20% of the total variance between sample means ($F(2,18)=1641.51$, $p<0.0001$) (**Figure 3.4.8B**). These data showed that the choice of cell extraction method and/or cryopreservation did not influence the percentage of CFU sub-types.

Similarly, viral transduction did not contribute to differences in CFU counts. In fact, there was not enough statistical power to assume that any of the tested blood processing conditions would favour the survival and/or proliferation of particular CFU types. The factor that was statistically proven to affect the percentage of CFUs was inherent to CFC progenitors and linked to their lineage-commitment fate. Of note, the majority of colonies were early erythroid-committed BFU-Es, only few of CFU-G/M/GM and none of CFU-GEMM. Lastly, we concluded that blood processing and/or transduction affected the frequencies of progenitor cells overall and not for a particular CFU type.

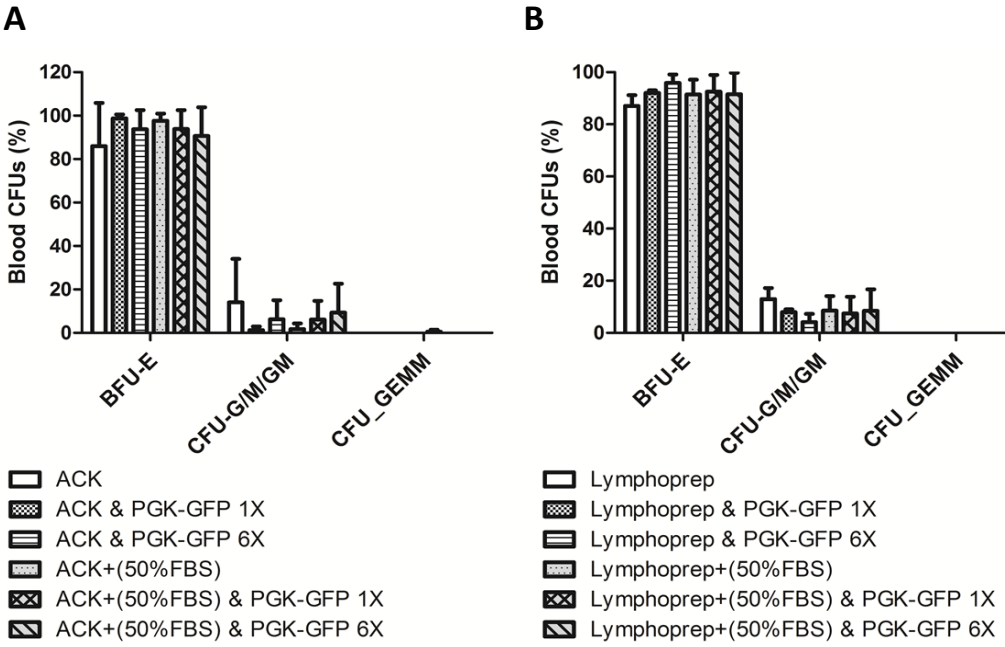


Figure 3.4.8. The effect of LV transduction on the percentage of CFUs for different blood processing conditions (cell extraction and/or freezing and/or transduction). Data represents the percentage (%) of CFU colonies BFU-E, CFU-G/M/GM and CFU/GEMM for **A)** ACK processed samples and **B)** Lymphoprep processed samples. Histograms show means ± SD. Statistical analysis was performed by two-way ANOVA.

We have also shown that the process of transduction, irrespective of the viral load, did not favour CFC differentiation towards a particular CFU colony type. We next investigated the efficiency of gene transfer using a large and a low viral volume in fresh or frozen cells separated by Lymphoprep or ACK. For VCN detection, direct detection of integrated vector copies in gDNA by qPCR is the method of choice. However, the small number of cells per CFU colony would have resulted in low DNA yield and rendered quantification by qPCR unreliable. Thus we confirmated LV-mediated gene transfer in CFCs indirectly, by scoring CFUs for GFP fluorescence. Indeed, it was to expedite this type of analysis that we originally chose to use the PGK-GFP vector to transduce the cells under test allowing detection of any progenitor of the myeloid lineage. Following CFU colony scoring and quantification, GFP-expressing CFUs (GFP-CFUs) were enumerated, and their percentage of the total CFU count per sample calculated. The raw data for two independent experiments is shown in **Figure 3.4.9A**. Expectedly, there was no detection of GFP-expressing CFUs in the non-transduced samples. However, the two experiments exhibited a somewhat large difference in scores (% CFU) for any given blood processing condition. Therefore, the percentage values were log transformed (log_{10}) to reduce residual variance, and mean values were analysed by two-way ANOVA. We showed that the choice of cell extraction method and/or cryopreservation did not influence the percentage of GFP-CFUs significantly ($F(3,12)=2.86$, $p=0.0815$). Instead, our results showed that transduction significantly affected the percentage of GFP-CFUs, accounting for 71.27% of the total variance between sample means ($F(2,12)=33.04$, $p<0.0001$) (**Figure 3.4.9B**)**.** The Bonferroni's post-hoc test for multiple comparisons identified a significant difference in the percentage of GFP-CFUs between transduced and non-transduced samples for all blood processing conditions with exception of fresh ACK preparations. For this latter condition, however, the number of GFP-CFUs was rather small (n=0.0±0.0, 6.44±4.7 and 14.9±18.1 for mock, 1x and 6x, respectively) and statistical power therefore reduced. Unsurprisingly, transduction increased the percentage of GFP-positive CFUs compared to mock samples. In addition we did not detect any difference in the percentage of GFP-CFUs with respect to the volume of vector added. These data showed an efficient and similar gene transfer irrespective of using a large or a low vector quantity. Of note, since the PGK promoter has been shown to be subject to silencing over time post-transduction [356], in all likelihood not all proviral integrations result in stable transgene

expression, owing to the suppressing effect of surrounding chromatin structure and DNA methylation [423].

We next investigated whether the transduction of GFP-CFUs was generic or directed towards a particular progenitor type. Analysis was performed on the log-transformed percentage of each CFU type using two-way ANOVA. As CFU-GEMM scores were zero for all samples, this CFU type was excluded from the analyses. Our results showed that for Lymphoprep-isolated leukocytes, the transduction of fresh or frozen preparations had a significant effect on the percentage of GFP-CFUs, accounting for 62.77% of the total variance ($F(5,12)=5.05$, $p=0.0101$). We did not detect a significant contribution by the differentiation fate of progenitors to differences observed in the percentage of GFP-CFUs ($F(1,12)=1.12$, $p=0.3108$) (**Figure 3.4.10A**). Expectedly, the Bonferroni's post-hoc test for multiple comparisons identified a significant difference in the percentage of GFP-CFUs for both BFU-E and CFU-G/M/GM types between the transduced (for both PGK-GFP X1 and X6) and non-transduced samples for both fresh and frozen Lymphoprep preparations ($p<0.05$). Analysis did not detect significant differences between fresh and frozen transduced samples (X1 vs. X6) suggesting that the effect of transduction on the percentage of GFP-CFUs was not influenced by the sample's condition (fresh or cryopreserved). Also, transduction was successful using either low or high vector quantities, with an observed, albeit not significant, increase in GFP-CFUs with the higher vector quantities. Compared to a low vector quantity, a high vector quantity resulted in increased GFP-CFUs by 5.8% (BFU-Es) and 28.6% (CFU-G/M/GM) for fresh preparations, and 11.2% (BFU-Es) and 185% (CFU-G/M/GM) for frozen preparations.

Similarly, our data showed that for ACK-isolated leukocytes transduction of fresh or frozen preparations had a significant effect on the percentage of GFP-CFUs accounting for 50.39% of the total variance ($F(5,12)=3.57$, $p=0.0329$). We did not detect a significant contribution by the differentiation fate of progenitors to differences observed in the percentage of GFP-CFUs ($F(1,12)=3.08$, $p=0.1046$) (**Figure 3.4.10B**). A Bonferroni's post-hoc test for multiple comparisons identified a significant difference in the percentage of GFP-expressing BFU-Es between the transduced (PGK-GFP X6) and non-transduced samples for the frozen ACK preparation ($p<0.05$). Even though not significant when compared to non-transduced control, all

other blood processing conditions resulted in GFP-expressing BFU-Es. Outcomes were very similar to those obtained with cells isolated using the Lymphoprep procedure. Although there was no significance for the effect of transduction, cell extraction and/or cryopreservation on the percentage of GFP-CFUs, clear trends were discernible from the data. For example, transducing with a 6-fold greater vector quantity resulted in an increase in the percentage of GFP-expressing BFU-Es by 146% in fresh preparations and by 272% in frozen preparations. However, because of high variance this difference between low and high vector quantities was not significant. In addition, GFP-expressing CFU-G/M/GM colonies were not detected in transduced (PGK-GFP X1) fresh preparations, either because these were not transduced or they were erroneously excluded owing to a weak fluorescence signal mistaken for auto-fluorescence. Overall, these data showed that the transduction was successful using either low or high vector quantities, with an observed increase in GFP-CFUs with the higher vector quantities. Lastly, blood processing using Lymphoprep density-gradient was somewhat superior to ACK in the transduction of a larger percentage of CFUs at both low and high vector quantities when compared to mock samples.

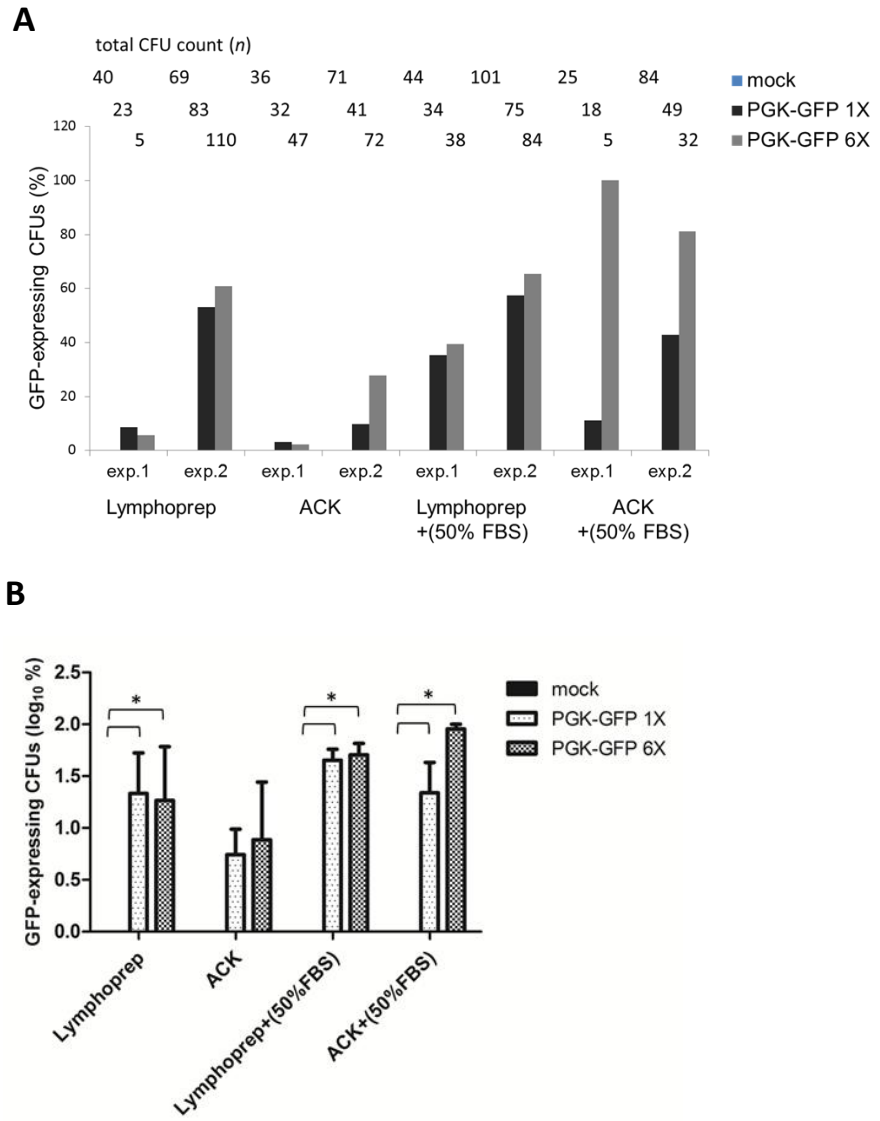


Figure 3.4.9. The effect of transduction on the proportion of GFP-expressing CFUs for different processing conditions (cell extraction and/or freezing and/or transduction). A) Histograms of absolute percentages of GFP-expressing CFUs from two independent experiments (exp.1 and exp.2). The percentages are calculated in relation to the total number of CFU colonies quantified per plate, and shown on top of each bar. **B)** Histograms of means ± SD of log-transformed percentages for GFP-expressing CFUs. Statistical analysis was performed by two-way ANOVA. Statistically different values are annotated (horizontal brackets & *, $p<0.05$).

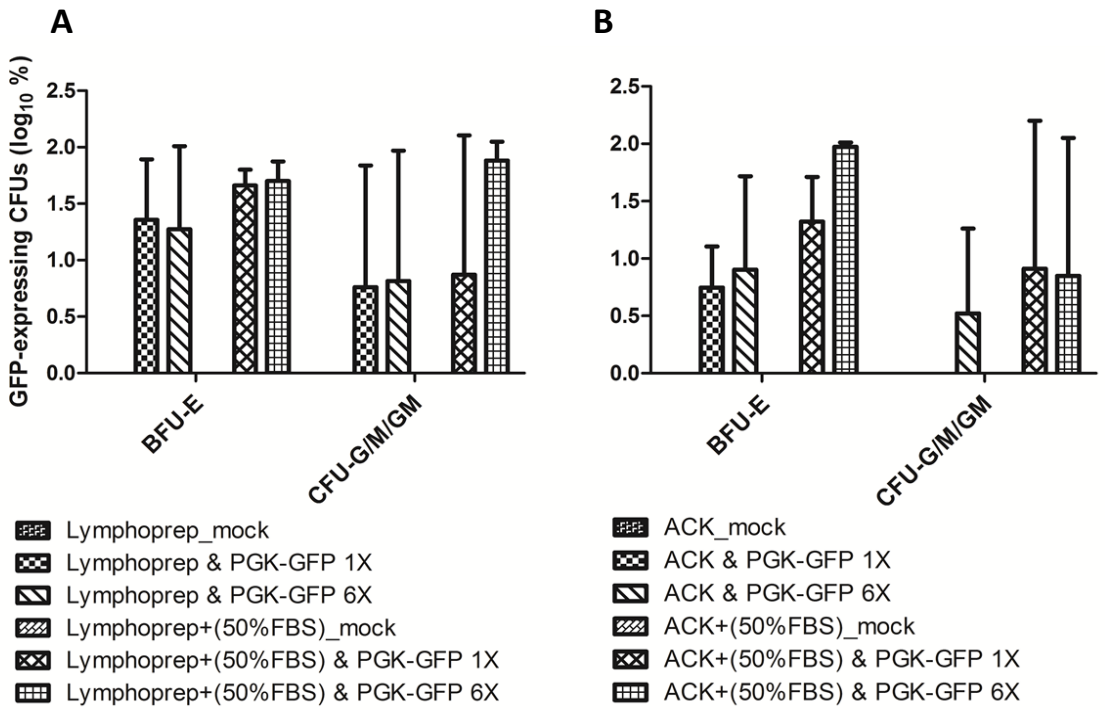


Figure 3.4.10. The effect of transduction on the proportion of GFP-expressing BFU-E and CFU-G/M/GM for different processing conditions (cell extraction and/or freezing and/or transduction). Histograms represent means ± SD of log-transformed percentages for GFP-expressing CFUs for **A)** Lymphoprep processed samples and **B)** ACK processed samples. Statistical analysis was performed by two-way ANOVA.

3.4.8 Discussion

The aim of the study presented in this chapter was the optimization of a protocol for the extraction and cryopreservation of leukocytes from small-volume PB samples for the establishment of a biorepository. The effectiveness of the protocol was evaluated by testing samples for leukocyte yield and viability, and for enumeration and evaluation of progenitor cells as CFUs in a semi-solid culture system. The protocol would be approved if cryopreserved samples shared similar characteristics to fresh samples.

The successful cryopreservation of whole blood as a source of lymphocytes has been reported in many epidemiological studies with minimal effect on cell integrity, viability, and DNA yield [424, 425]. In our study, initially whole blood was frozen using a mixture of 90% FBS (*Life Technologies*) with 10% DMSO for inexpensive large-scale storage. The extraction of leukocytes from frozen blood was performed using Ficoll Paque PLUS, a widely used and highly robust density-gradient purification reagent. Preliminary analyses showed ~90% reduction in PBMC counts with the remaining cells lacking haematopoietic potential. The protocol was therefore dismissed for our purposes.

Blood was then processed using Ficoll Paque PLUS density-gradient, and separated PBMCs were cryopreserved in 90% FBS-based freezing media. Analysis showed that the post-thawing viable cell count was reduced by 32% compared to fresh preparations. Studies have shown that the length of time a blood sample is stored before processing can influence PBMC recovery after thawing, and suggested that a 24 h storage would result in decreased PBMC viability (~86%–92%) compared to processing within 8 h (≥94%) [408]. Freezing would also affect PBMC viability, the extent of which was shown to vary with the method of cryopreservation [426]. In our investigation, blood was processed within 24–30 h of collection, as these blood samples were collected for testing prior to RBC transfusion and only shipped to our facilities the morning after.

The density-gradient technique is a labour-intensive process requiring approximately 50–60 minutes for each sample. When processing many small samples simultaneously, although feasible for experienced personnel and saving time compared to individual processing, sample

preparation time is still increased further. In order to reduce costs and save time, we attempted a simple whole blood lysis method, which uses a self-made ACK lysing buffer. Briefly, the ACK lysis solution consists of ammonium chloride (NH_4Cl), which accumulates in RBCs and causes them to rupture via osmotic pressure. Studies comparing different commercial RBC lysis buffers indicated that the ACK solution had a small effect on leukocyte viability [405, 406]. In our study, the lysis protocol is comparatively straight-forward and completed within approximately 30 minutes even for multiple samples. Given that the time is critical for optimal lysing, the number of samples processed simultaneously is mainly dependent on the celerity and dexterity of the technician. In order to enable comparisons to the Ficoll-Paque PLUS density-gradient technique, the ACK-isolated NCs were cryopreserved in 90% FBS-based freezing media. Analysis showed a superior performance using ACK lysis rather than Ficoll-Paque PLUS density-gradient separation in terms of the frequencies of haematopoietic progenitors, with the former exhibiting a 1381% greater progenitor output following thawing. Since only two frozen samples of ACK-isolated NCs were analysed, additional experiments should be set-up to confirm this observation.

Recently, we have established a two-phase liquid culture system (protocol C, see **section 3.5.2.3**) for the expansion and differentiation of $CD34^+$ cells. Characteristic features of this protocol include the extraction of PBMCs using Lymphoprep^{TM} density-gradient medium and the cryopreservation of $CD34^+$ cells with a mixture of 50% FBS (*Hyclone*) and 10% DMSO. Our aim is to use protocol C as a routine method for the expansion of $CD34^+$ progenitors in frozen leukocyte samples. Therefore we tested the Lymphoprep density-gradient medium for PBMC separation, followed by cryopreservation with 50% FBS (*Hyclone*). As the serum was very expensive, both in terms of its commercial brand as well as the volume added per sample, we tested the cryopreservation of PBMCs with 30% FBS (*Life Technologies*), which was used in the cryopreservation of other mammalian cells. Our study compared the two density gradient media – the Ficoll-Paque^{TM} PLUS and the Lymphoprep^{TM}. Although there was no difference in the number of viable leukocytes in fresh and/or frozen PBMCs separated by either Ficoll-Paque PLUS or Lymphoprep, analysis revealed a significant variation in progenitor frequencies after thawing. The best of the density-gradient media, according to the protocols we used, was

Lymphoprep since the cryopreservation of PBMCs separated by Ficoll-Paque PLUS resulted in a 90% loss of progenitors. Our results recommend the use of a 50% FBS freezing medium for the cryopreservation of PBMCs separated by Lymphoprep.

Furthermore, in order to enable comparisons to the Lymphoprep density-gradient technique, the ACK-isolated NCs were similarly cryopreserved in 50% FBS (*Hyclone*) and 30% FBS (*Life Technologies*) freezing media. Analysis showed increasing losses of viable leukocytes with a lower FBS concentration in the freezing medium. In fact, both 50% and 30% freezing media resulted in 90%–95% loss of progenitors. In contrast, cryopreservation using a 90% FBS (*Life Technologies*) freezing medium did not negatively affect viable leukocyte counts or their constituting progenitor frequencies, and thus appeared most suitable for the cryopreservation of ACK-isolated leukocytes.

Lastly, we compared the effect of the ACK+(90% FBS) and the Lymphoprep+(50% FBS) blood processing conditions on post-thaw progenitor frequencies. Our results showed that the former protocol had a better performance, resulting in 55% higher progenitor frequencies. As a final experiment, we investigated the transducibility of fresh and frozen leukocyte fractions. We found that fresh and frozen PBMCs separated by Lymphoprep density-gradient were equally permissive to viral transduction. We did not detect any changes in progenitor frequencies or bias towards a particular CFU colony type. Transducibility experiments were set-up to compare gene transfer efficiency between leukocyte fractions of Lymphoprep and ACK extractions. At the time, the frozen preparations transduced for both extraction protocols had been cryopreserved with a 50% FBS freezing medium. The decision was made on the basis that a 50% FBS-based freezing was suggested as being suitable in the erythroid liquid culture protocol, which we had set out to establish and use. However, with the new data in hand, transduction experiments should also be conducted with ACK-isolated leukocytes cryopreserved in 90% FBS-based freezing media. Overall, more samples per combination of procedures are required for a reliable comparison of protocols.

As a result of these analyses and pending further data, we chose the Lymphoprep density-gradient in combination with a 50% FBS-based freezing medium for the creation of a biobank

based on PBMCs of different β-thalassaemia genotypes. With β-thalassaemia patients receiving periodic blood transfusions as supportive therapy, we expect to receive small-volume pre-transfusion-test blood samples from any given patient every 2–3 weeks. To date, our cellular biobank holds 117 thalassaemia samples of the following genotypes: IVSI-110/IVSI-110, IVSI-110/IVSI-1, IVSI-110/IVSI-6, IVSI-6/IVSII-745, and IVSI-110/C39 [Note: IVSI-110: HBB:c.93-21G>A; IVSI-1: HBB:c.93-21G>A/T/C; IVSI-6: HBBc.92+6T>C; IVSII-754: HBB:c.316-106C>G; C39: HBBc.118C>T]. Within the context of this project, this protocol enables the collection and pooling of PBMCs from up to 30 mL fresh blood, which is required to set-up one erythroid liquid culture and which is difficult to obtain from β-thalassaemia patients on a regular volunteer basis. In contrast to the two-phase liquid culture protocol C, the protocol we propose herein does not use immunoaffinity magnetic bead-based $CD34^+$ cell isolation since it is a costly technique, application of which on many small-volume blood samples would render the whole procedure impractical. Using selection for $CD34^+$ cells will eliminate a lot of the more committed ($CD34^-$) erythroid progenitor cells towards a more homogeneous cell population. However, it has been shown that these $CD34^-$ cells would make a substantial contribution to erythroid expansion when PBMCs are used as a starting material for *in vitro* liquid cultures [427]. Essentially, cryostorage of same-patient cells over time will enable large-scale experiments whereby pooled samples are thawed and cultured in liquid media, which support the expansion of haematopoietic progenitors.

3.5 Functional characterization of lentiviral shRNA vectors in primary human erythroid cell cultures

3.5.0 Aims

- To establish optimal and reproducible *in vitro* culture conditions for large-scale expansion of progeny from human $CD34^+$ cells.
- To characterize expression of vector-derived *HBB* and concomitant induction of endogenous *HBG* from shRNA- and shRNA-miR-encoding MA821Q-derived vectors in erythroid progenitors from normal and β-thalassaemia patients.

3.5.1 Introduction

$CD34^+$ cells are an attractive target for gene therapy (GT) and instrumental to the advancing field of autologous haematopoietic stem and progenitor cell (HSPC) transplantation. Standard GT clinical protocols incorporate *ex vivo* gene transfer procedures using technologies which allow the isolation of $CD34^+$ cells from patients, their expansion and genetic modification *in vitro*, and their infusion back to the patient to establish a graft of gene-corrected cells. The therapy for haematological diseases depends on regeneration of short-lived cells and would therefore benefit from the ability of $CD34^+$ cells to self-renew and have a longer life-span. In this case, the engraftment of genetically modified stem cells will generate a functional progeny that reconstitutes the BM niche and offers a long-term correction of the disease [163, 428].

In vitro genetic modification of $CD34^+$ cells is based on functional transduction protocols and optimized culture conditions. At first, GT technology for β-thalassaemia employed γ-retroviral vectors (RVs) for gene delivery to HSPCs. Initial experiments reported low gene transfer efficiency since adult stem cells are maintained in a quiescent state and RVs would selectively transduce mitotic cycling cells. Consequently, protocols were modified to include pre-stimulation

of stem cells with cytokines in order to induce cell cycle entry and proliferation. Unfortunately, culture conditions that promoted proliferation were in most cases associated with an irreversible change of stem cell phenotype, that is, loss of self-renewal and ability to reconstitute haematopoiesis after transplantation, which in turn compromized therapy. This led to the development of improved protocols, which would allow efficient gene transfer in the presence of an optimal cytokine composition formulated to achieve prolonged expansion of $CD34^+$ cells in culture [163, 429, 430]. Therapy requires an efficient and stable transduction of most of the $CD34^+$ cells harvested for transplantation. The use of self-inactivating HIV-based lentiviral vectors (LVs) presented a promising alternative to RVs owing to their ability to infect and replicate in both proliferating and non-proliferating cells [431, 432]. Others have reported efficient LV gene transfer (15%–25%) in cord-blood NOD/SCID-repopulating HSCs in the absence of cytokine stimulation, which was, nevertheless, significantly enhanced (by ~2-fold) with additional short cytokine exposure [433]. Further work identified that proteasome activity in transduced $CD34^+$ cells would impede viral integration and could be down-regulated rapidly and substantially by cytokine pre-stimulation, increasing permissiveness of $CD34^+$ cells to LV gene transfer [434]. Even though both LV- and RV-based transduction protocols depend on cytokine stimulation, the former system would achieve a more efficient transduction with shorter culture times, thus lowering the risk of losing stem cell properties [163, 433, 435, 436].

As a pre-requisite to clinical trials, the therapeutic safety and efficiency of vector-mediated gene transfer in $CD34^+$ cells is assessed in suitable animal models for systemic and long-term analysis on the one hand, and on the other *in vitro* in primary cells (usually $CD34^+$ HSPCs) of healthy controls prior to assessment in patient-derived HSPCs. Over the years, protocols have been developed for reproducible culture and large-scale expansion of genetically and functionally normal HSPCs, which often need some modification for application to patient-derived cells [437, 438].

In the field of HBB gene therapy, vectors are screened *in vitro* for their ability to correct the phenotype of erythroid progenitors from patients preparing for gene therapy. *In vitro* assays comprise colony-forming assays in semi-solid culture medium and liquid suspension cultures. The first culture method represents a one-step culture in which HSPCs proliferate and

differentiate into discrete colonies in response to cytokine stimulation, which in turn allows a retrospective assessment of the HSPC population after gene transfer. This assay is a useful tool to detect and quantify treatment-related changes in progenitor frequencies and lineage composition [188, 439, 440]. However, haemoglobinized colonies will not reach terminal differentiation into mature erythroblasts [438], limiting phenotypic and biochemical analyses of this lineage. The second culture method is a two- or multi-step liquid culture in which cells are grown in the presence of different cytokines and other cellular regulators [429, 441, 442] to induce the expansion and differentiation of $CD34^+$ progenitors. This assay can be scaled-up for large-scale experiments. An optimal erythroid liquid culture system would result in i) substantial expansion of $CD34^+$ progenitors, ii) unilineage erythroid commitment and differentiation, and iii) complete terminal maturation into functional enucleated cells [443]. The need for an optimal liquid culture system led the way towards the development of a number of published protocols. Even though a significant expansion of erythroid progenitors was reported, the majority of protocols would result in weak or moderate (10%–50%) erythroid enucleation [403, 439, 440, 444-446]. Further advances have recently enabled both high expansion and high enucleation efficiency of cultured $CD34^+$ cells. Best yields to date have been reported by Giarratana et al [443] with $1.95x10^6$-fold expansion and 100% enucleation, by Miharada et al [447] with $7.2x10^5$-fold expansion and 100% enucleation, and by Fujimi et al [448] with $3.5x10^6$-fold expansion and 99.4% enucleation, all three from CB-derived $CD34^+$ cells. An in depth description of the great density of *in vitro* erythropoietic culture systems is beyond the scope of this thesis, with differences, e.g., in cell source, number of growth phases, use of cytokines and reliance on or independence from co-cultures impeding a concise comparison. My work does not involve the development of novel culture systems or the modification of extant protocols in order to optimize cultivation conditions, which are lengthy and costly research projects on their own right. Simply, this project will use established liquid culture systems, adapted to locally available reagents from detailed protocols of collaborators, to demonstrate the efficacy of vector-mediated gene transfer.

3.5.2 Erythroid liquid culture systems

3.5.2.1 Erythroid liquid culture – Protocol A

The erythroid culture system described by Migliaccio et al [335] is a two-phase method for production of erythroid cells from either light-density cells or $CD34^+$ cells purified from blood. It consists of the proliferation-phase and the differentiation-phase cell cultures. A modified version of this method was described by Roselli et al [188] for characterization of the GLOBE vector (MA821) in progenitor cells of β-thalassaemia patients.

Herein, $CD34^+$ progenitor cells from peripheral blood (PB) or buffy coat of healthy individuals were grown in Phase-I media in the presence of erythropoietin (EPO), SCF, IL3, dexamethasone, and β-estradiol. The haematopoietic cytokine SCF was reported to stimulate the growth of haematopoietic progenitor cells, while dexamethasone and β-estradiol were shown to sustain their proliferation. SCF, in the presence of growth factor IL3, was shown to stimulate erythroid differentiation. Similarly, EPO on its own was reported to shift the propensity of progenitor cells towards differentiation [442, 449, 450]. Initial optimization experiments used the cytokine combination described by Migliaccio et al [335]. Using these conditions, we reported good differentiation during the proliferation-phase such that the differentiation-phase was not included in our liquid method. The proposed scheme of the procedure is shown in **Figure 3.5.1** below:

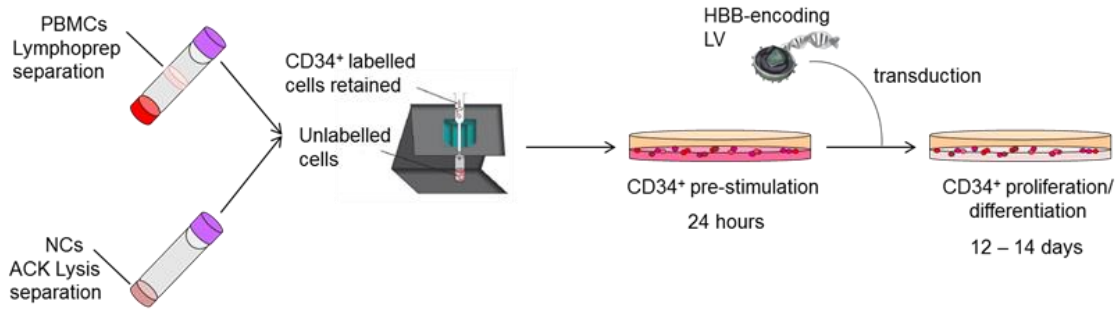
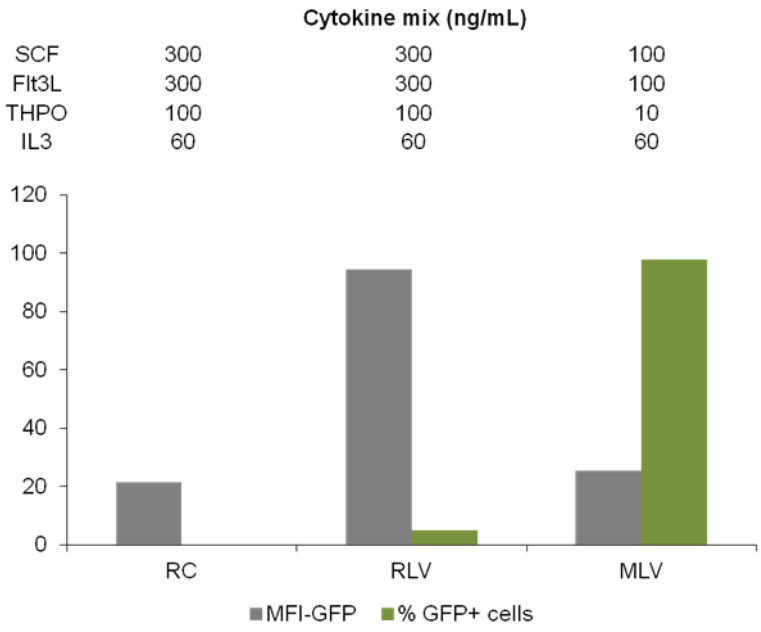


Figure 3.5.1. Protocol A. Outline of the cell culture procedure adapted from Migliaccio et al [335]. $CD34^+$ cells purified from PB are stimulated with cytokines for 24 h prior to transduction at various MOIs, with or without retronectin. After transduction, cells are maintained in liquid culture and analysed 12–14 days later.

LV transduction of HSPCs is effective in the absence of cell proliferation; however, cytokine pre-stimulation to prompt cell-cycle progression is generally performed to improve susceptibility of HSPCs to transduction. For this study, we employed a highly efficient transduction protocol described by Roselli et al [188]. Preliminary PCR gel analysis for viral sequences in transduced bulk $CD34^+$ samples resulted in detection of only faint bands for the virus-specific product (data not shown), which prompted a review of the transduction protocol. The investigation compared between the Roselli protocol and an alternative transduction protocol described by Millington et al [451], which uses the same cytokine combination but at different concentrations. Transduction at a MOI of 25 was performed on retronectin-coated 24-well plates (5 μg/cm^2) using the PGK-GFP vector for constitutive *GFP* expression. Transduction efficiency was estimated as the percentage of GFP^+ cells analysed using flow cytometry. Following pre-stimulation with the *Roselli* cytokine protocol, only a small proportion of cells (~5%) were GFP^+, exhibiting an ~339% increase in mean fluorescence intensity (MFI) compared to non-transduced cells (when gating for 4.6% false GFP^+ cells with MFI=21.05 in the whole control-cell population). These efficiency and MFI data suggested a biased transduction of a small number of cells, possibly with multiple integrants per cell. Conversely, pre-stimulation with the *Millington* cytokine protocol resulted in a large number of GFP^+ cells (~98% cells) with a moderate ~18% increase in MFI-GFP compared to non-transduced cells. These data in turn suggested a uniform transduction of almost all cells, possibly with a single integrant per cell. Overall, my initial assessment of these two similar transduction protocols showed that under the conditions and with the reagents employed both performed quite differently, that the *Millington* protocol gave negligible transgene expression unsuitable for further analyses and that the *Roselli* protocol, reported as highly efficient elsewhere, would not be suitable for my purposes without optimisation and an increase in transduction efficiency (**Figure 3.5.2**).



	MFI-GFP		GFP⁺ cells
	MFI	Δ% (fold change)	% +ves
RC	21.5		0
RLV	94.43	339.21 (4.39)	5.02
MLV	25.32	17.77 (1.18)	97.67

Figure 3.5.2. Transduction efficiency of CD34$^+$ cell in CellGro medium supplemented with various cytokine concentrations. CD34$^+$ blood cells from normal donors were pre-stimulated for 24 h in various cytokine combinations (ng/mL) as indicated and transduced with the PGK-GFP vector at a MOI of 25 in retronectin-coated plates. The histogram represents values of GFP mean fluorescence intensity (MFI-GFP) and percentage (%) of GFP$^+$ cells determined by flow cytometry after 1 day in culture (2 days post-transduction). The table shows MFI-GFP, the percentage change (Δ%) and fold change in MFI-GFP between the transduced (RLV, MLV) and the control (RC) cultures. Data represent single values from one experiment. RC: Roselli protocol+control; RLV: Roselli protocol+LV; MLV: Millington protocol+LV.

The optimization procedure included the use of different tissue-culture plates to enhance retronectin-coating efficiency and using different dilutions with new cytokine solutions. In the experiment presented here, transduction of CD34$^+$-enriched cells was performed on retronectin-coated *Corning* 35-cm^2 plates (5 µg retronectin/cm^2) using the PGK-GFP vector at a MOI of 100. The transduction efficiency was analysed for i) vector-encoded mRNA by RT-qPCR, and ii) integrated VCN by qPCR. Transduction was performed 24 h post-stimulation and analysis was on day 13 during the proliferation-phase of culture. All cultures exhibited ≥90% positivity for

benzidine staining at a relatively low toxicity (cell death 17.4±3.22%, n=4), indicating that the culture method supported the differentiation and accumulation of Hb-producing erythroid cells. In addition, the proliferation-phase achieved proliferation of cells to 59.5-times the initial inoculum of $4.0x10^4$ $CD34^+$ cells ($2.38±0.5x10^6$ cells, n=4). Endpoint analysis of differentiation was assessed by May-Grünwald-Giemsa staining on portions of the cultured cells to determine morphology. Representative cytocentrifugation samples are shown in **Figure 3.5.3**. The majority of cells (n=2, scoring a total of 183 cells) consisted of orthochromatophilic normoblasts (81.32±3.18%) with few polychromatophilic normoblasts (5.63±1.90%), basophilic normoblasts (1.88±0.63%), proerythroblasts (2.23±0.13%), and enucleated erythrocytes (1.88±0.63%). Previous attempts to re-culture cells in differentiation-phase media stimulated with EPO and insulin had resulted in gross cell loss (data not shown). Relative transduction efficiency and vector-derived expression were estimated by quantification of GFP mRNA levels using RT-qPCR. Transduction of cells pre-stimulated with the *Roselli* cytokine protocol resulted in ~57-times the GFP mRNA levels (0.997) observed as background in non-transduced cells (0.017) with an average VCN of 22 in bulk samples. Transduction of cells pre-stimulated with the *Millington* cytokine protocol resulted in GFP mRNA levels (0.927) ~68-times those of controls (0.014), with a comparable average VCN of 20.7 in bulk samples (**Figure 3.5.4**). Although these findings were of limited authority without repetition, the differences detected for both protocols were negligible. While the *Roselli* protocol relied on higher cytokine concentrations (**Figure 3.5.2**) and was therefore costlier, it had been confirmed to work for the same transduction and differentiation procedure (albeit based on BM-derived HSPCs) [188], so that we continued to use it for pre-stimulation. .

$CD34^+$ cells were isolated from the leukocyte fraction of PB from healthy individuals by positive selection using a magnetic cell separator. The purity of $CD34^+$-enriched cells was then analysed by flow cytometry using a PE-conjugated anti-CD34 antibody. However, the analysis repeatedly produced low fluorescence intensity and did not give evaluable enrichment data (data not shown). As the MACS technology by *Miltenyi Biotecy* is among the most reliable methods for $CD34^+$ cell enrichment, with >90% purity [439, 440, 445], we speculated that the anti-CD34 antibody or staining protocol were probably not functional. In support of this speculation, we

noted during endpoint analysis the level of cell proliferation (see above) and that nearly all of the cells in culture consisted of erythroid cells (**Figure 3.5.3)**, indicative of the progenitor and haematopoietic phenotype of the starting cell population.

The success of the optimized erythroid liquid-culture protocol was mainly limited by the small number of $CD34^+$ cells obtained from PB for large-scale experiments with all vectors. Specifically, MACS-selection of $CD34^+$ cells for NCs isolated from fresh blood (30–52 mL) by a simple ACK lysis procedure ($0.66–4.79x10^7$ cells/mL, n=6) resulted in a low yield of only 0.4±0.2% of selected cells ($0.20–1.72x10^5$ cells/mL, n=6), giving a total of $0.92–3.44x10^5$ $CD34^+$-enriched cells (**Table 3.5.1)**. Assuming that the frequency of $CD34^+$ cells in PB leukocyte fraction is 0.03%–0.09% [452], a 4–10-fold enrichment of $CD34^+$ cells was achieved by using the MACS technique. We reported that moderate volumes (30 mL) of PB from healthy donors (n=3) would result in ~$1.03±0.1x10^5$ total $CD34^+$ cells (**Table 3.5.1**). The cultures shown in **Table 3.5.2** were each started with a different number of $CD34^+$ cells mainly owing to variability in the number of $CD34^+$-enriched cells depending on the individual blood donor from whom the cells had been derived. The intrinsic properties of the cells used to start the cultures, as well as the culture conditions, appear to influence the level of amplification of each culture thus resulting in widely variable of cell amplification overall. Expansion was always low, and viability was occasionally compromised, especially with a starting sample of transduced $CD34^+$-enriched cells, so that several cultures were discarded without further analysis (data not shown). Even though the proliferation-phase of cell culture supported cell growth and amplification, it also induced erythroid differentiation early-on, so that by day 13 more than 50% of cells were haemoglobin (Hb) positive (**Table 3.5.2**). Since differentiated cells are not optimal targets for GT, the liquid-culture protocol was not used for $CD34^+$ cell expansion.

Starting with a small number of $CD34^+$ cells (≤$5x10^5$), the optimized liquid culture protocol produced enough cell material for the quantification of globin mRNA by RT-qPCR, but not for globin protein analysis. Moreover, VCN measurements by qPCR appeared too high for the respective MOI used and in comparison to transduction efficiencies reported by others. The working assumption was that the low cell number available for each experiment, either by giving too little, too dilute or too impure sample material for different assays, was at the root of these

apparent discrepancies. Data of one representative experiment (**Table 3.5.2, exp. 4**) are presented in **Figure 3.5.5**. Owing to the prohibitive cost of retronectin solution, the transduction procedure was modified to accommodate simpler and less expensive practices such as those described in **section 2.9.5.3.3**. Enriched-$CD34^+$ cell populations ($0.33x10^5$) were transduced with the MA821Qsh_B9 and the MA821Q vectors using an MOI of 40 and cultured in proliferation-phase medium for 12 days. The timing of cell harvest was determined by quantifying haemoglobinization using benzidine staining. In order to ensure consistency at the level of erythroid differentiation between samples, cultures were collected when >90% of cells scored positive for benzidine staining. Cell expansion was recorded at 188% for the non-transduced sample, and at 415% and 203% for the MA821Qsh_B9 and MA821Q transductions, respectively. In spite of an efficient proliferation and differentiation, the transduced samples exhibited (2.48±1.21)-fold (8%–17%) higher cell death compared to the non-transduced sample (5%). Previous experiments indicated massive cell loss in prolonged culture conditions, especially if cultures exhibited high cell expansion. The need for a large cell sample to ensure sufficient material for molecular analysis prompted the collection of cells by day 12. The efficiency of shRNA-mediated knockdown of BCL11A transcripts was assayed by RT-qPCR, (**Figure 3.5.5)**. Compared to the non-transduced sample, the MA821Qsh_B9 vector resulted in a 34% reduction of BCL11A mRNA accompanied by a similar reduction (~30%) in HBG mRNA with an average VCN of 61.7. Surprisingly, cells transduced with the MA821Q vector, a negative control for shRNA-mediated BCL11A knockdown, still showed a 23% reduction in BCL11A mRNA accompanied by an increase (~11%) in HBG mRNA with average VCN of 90.7. At face value this observation is not of any biological significance, since knockdown of *BCL11A* expression would relieve repression on *HBG* expression, resulting in an increase of HBG mRNA in adult erythroid progenitor cells. At the time the optimization of the liquid-culture protocol was in progress, many of the current RT-qPCR gene-specific assays were not established in our laboratory. BCL11A mRNA expression was quantified using a primer set which amplified regions 5' upstream of the shB9 and shB1 siRNA target sequences, instead of amplifying regions encompassing the siRNA recognition sites. Also, the T87Qx2x3 primer pair for quantification of vector-derived HBB was not yet standardized. The available primer pair, specific to both the endogenous and vector-derived HBB, produced primer dimers which several

modifications to annealing/extension temperatures failed to eliminate (data not shown). Therefore, this study did not provide information on *HBB* expression which, if detectable at higher levels than control, would indicate normal processing of *HBB*, without interefence by intronic shRNA sequences or the choice of intronic cloning site.

Overall, the optimized liquid culture system exhibited mass production of erythroid progenitors starting with a low number of $CD34^+$ cells. Even though the cultures that survived to the end were efficiently differentiated based on positive Hb staining, the resulting cell material was insufficient to perform all necessary analyses. The main limiting factor of this system was the small number of $CD34^+$ cells obtained from PB for large-scale experiments.

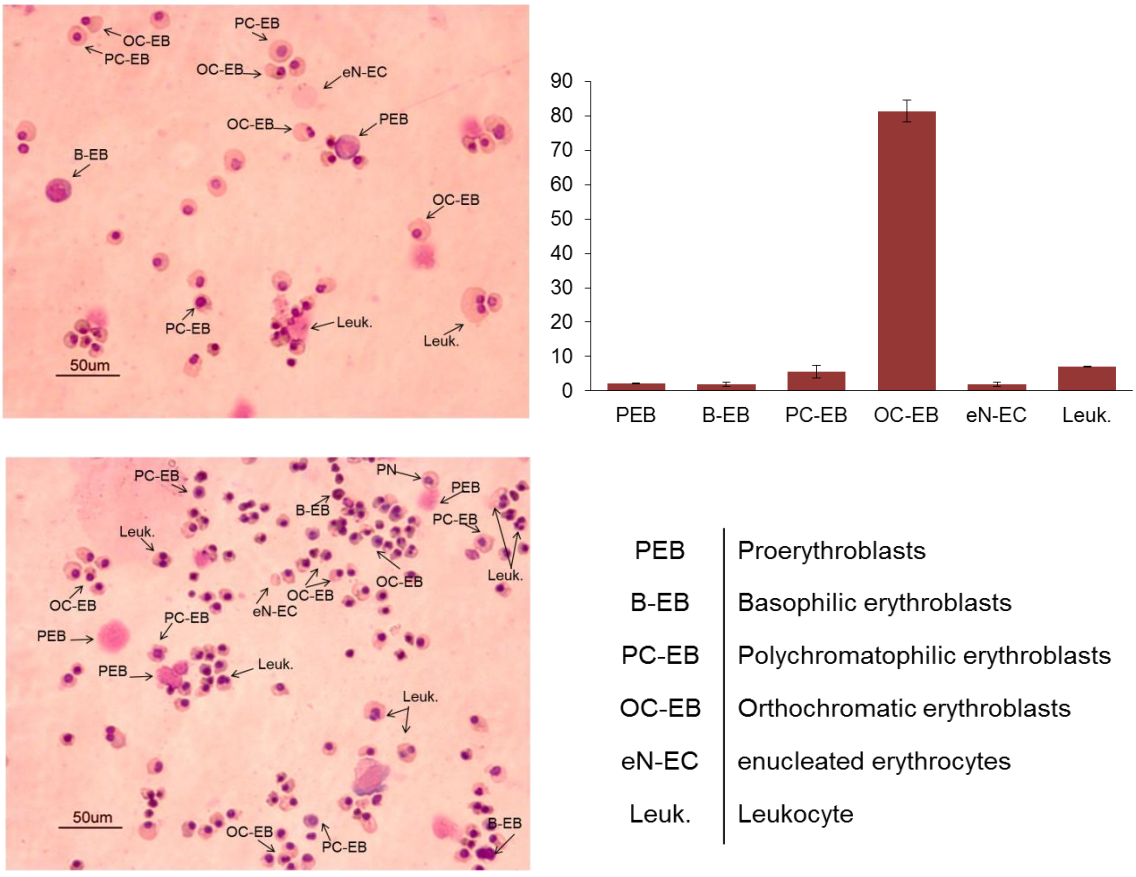
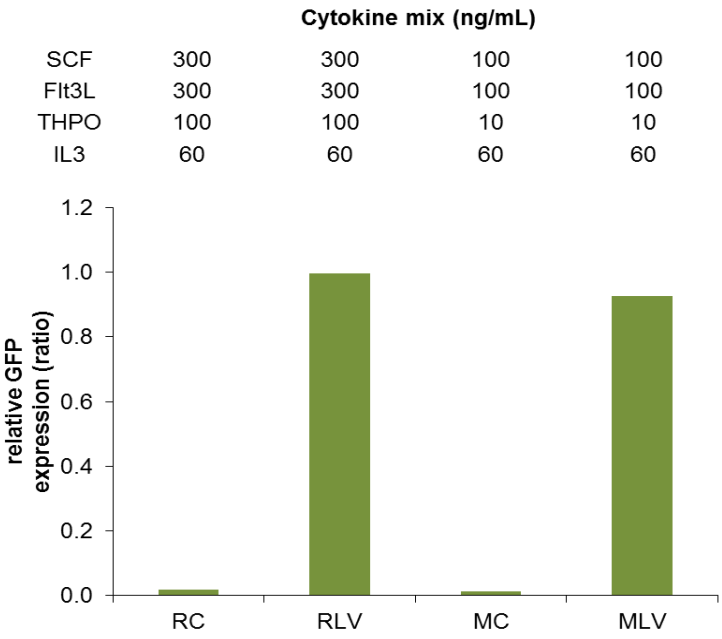


Figure 3.5.3. *In vitro* erythropoiesis starting from $CD34^+$ blood cells. Erythroid progenitors from normal donor cultures were subjected to cytocentrifugation and staining with May-Grünwald-Giemsa on day 13 and analysed for the proportion of proerythroblasts (PEB), basophilic erythroblasts (B-EB), polychromatophilic erythroblasts (PC-EB), orthochromatophilic erythroblasts (OC-EB), enuclated erythrocytes (eN-EC), and leukocytes (Leuk.). The histogram represents the mean ± SD percentage cell number for each erythroblast type from the two images. Size bar 50 µm.



R^2 = 0.967, *E* = 1.71

	relative GFP expression		VCN
	ratio	Δ% (fold change)	
RC	0.017		0.09
RLV	0.997	5601.39 (57.01)	22.0
MC	0.014		0.06
MLV	0.927	6710.34 (68.10)	20.7

Figure 3.5.4. Transduction efficiency of CD34$^+$ cells in CellGro medium supplemented with various cytokine concentrations. CD34$^+$ cells from normal donors were pre-stimulated for 24 h in various cytokine combinations (ng/mL) as indicated and transduced with the PGK-GFP LV at a MOI of 100 in retronectin-coated plates. Transduced cells were analysed for GFP mRNA levels by RT-qPCR after 14 days in culture. Values are normalized by *GAPDH* expression. The table shows relative *GFP* expression, the percentage change (Δ%) and fold change in expression between the transduced (RLV, MLV) and the control (RC, MC) cultures, as well as the average cell VCN in bulk samples. Data represent single values from one experiment. RC: Roselli protocol+control; RLV: Roselli protocol+LV; MC: *Millington* protocol+control; MLV: Millington protocol+LV; E: PCR amplification efficiency; R^2: square of Pearson's correlation coefficient.

blood (mL)	NCs ($x10^7$)/mL	NCs (total count; $x10^8$)	$CD34^+$-enriched cells ($x10^5$)/mL	$CD34^+$-enriched cells (total count; $x10^5$)	relative yield (%)
30	1.36	0.82	0.53	1.06	0.39
30	0.66	0.40	0.56	1.12	0.85
30	1.58	0.95	0.46	0.92	0.29
40	0.95	–	0.22	1.70	0.23
50	4.79	9.57	1.72	3.44	0.36
56	2.06	1.23	0.20	1.00	0.10

Table 3.5.1. Efficiency of enrichment of $CD34^+$ cells from ACK-isolated leukocytes of normal donor blood by magnetic-activated cell sorting (MACS). The table shows the number of ACK-isolated cells (NCs) and MACS-selected $CD34^+$ cells, as determined by trypan-blue staining. The proportion of $CD34^+$ cells in the NC population is calculated as the relative percentage (%) yield [($CD34^+$ cells/NCs)*100].

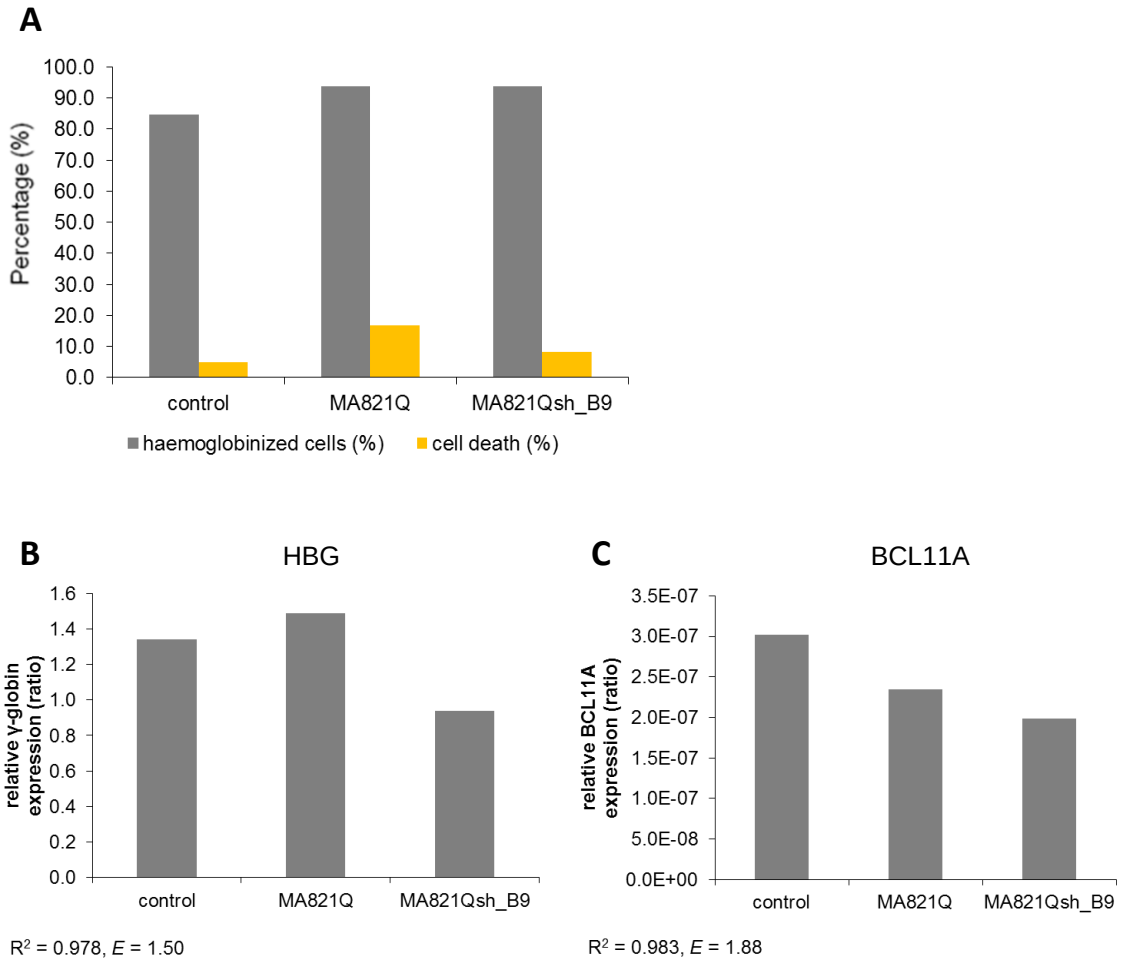
non-transduced cells

exp.	$CD34^+$ cell count ($x10^5$)		Δ%	cell death (%)	haemoglobinized cells (%)
	Day 0	Day 12 - 14			
1	0.49	0.40	-18.4	65.03	17.65
2	1.00	0.07	-93.0	93.00	25.00
3	0.20	1.52	+660.0	13.63	86.67
4	0.33	0.95	+187.9	5.00	84.62
5	0.24	0.95	+295.8	48.49	80.00
6	0.86	1.21 ($x10^6$)	+1306.9	29.46	61.84
7	1.15	0.24	-79.1	33.30	83.30

transduced cells

exp.	MOI	$CD34^+$ cell count ($x10^5$)		Δ%	cell death (%)	haemoglobinized cells (%)
		Day 0	Day 12 - 14			
1	100	0.49	0.53	+8.2	50.00	8.33
		0.49	0.35	-28.6	62.07	12.50
2	100	1.00	0.00			
3	100	0.45	2.16	+380.0	38.64	90.0
		0.45	2.40	+433.3	48.28	77.27
4	40	0.33	1.70	+415.1	8.10	93.75
		0.33	1.00	+203.0	16.67	93.75
5	10	0.24	0.62	+158.3	54.17	79.17
	40	0.24	5.04	+2000	50.00	78.57
	10	0.24	1.01	+320.8	41.94	72.00
	40	0.24	0.73	+204.2	45.83	75.00
6	40	0.80	1.69 ($x10^6$)	+2012.5	17.80	73.00
		0.80	3.53	+341.3	31.17	52.50
		0.80	2.73	+241.3	33.87	62.96
7	100	1.15	0.40	-65.2	64.30	50.00
		1.15	0.40	-65.2	23.10	66.70

Table 3.5.2. *In vitro* proliferation and erythroid differentiation of $CD34^+$ cells. $CD34^+$ cells from normal donors were pre-stimulated for 24 h prior to transduction with different LVs at various MOIs as indicated. Transduced cells were analysed by trypan-blue and benzidine staining after 12–14 days in culture. The tables show the percentage change (Δ%) between the initial (Day 0) and final (Day 12–14) cell numbers (increase (+); decrease (–)), as well as the percentage (%) of cell death and haemoglobinization for control (top table) and transduced (bottom table) cultures.



R^2 = 0.978, *E* = 1.50

R^2 = 0.983, *E* = 1.88

	relative gene expression				VCN
	HBG		BCL11A		
	ratio	Δ% (fold change)	ratio	Δ% (fold change)	
control	1.34		$3.02x10^{-07}$		0.08
MA821Q	1.49	11.02 (1.11)	$2.34x10^{-07}$	-22.30 (0.78)	90.66
MA821Qsh_B9	0.94	-29.95 (0.70)	$1.99x10^{-07}$	-34.12 (0.66)	61.66

Figure 3.5.5. BCL11A knockdown in primary erythroid cultures using shRNA-encoding LVs. $CD34^+$ cells from normal donors were pre-stimulated for 24 h prior to transduction with the MA821Q and MA821Qsh_B9 LVs at a MOI of 40. Cells were analysed for **A)** the percentage (%) of haemoglobinization (grey bars) and cell death (yellow bars) by benzidine and trypan blue staining, respectively, as well as for **B)** HBG and **C)** BCL11A mRNA levels by RT-qPCR after 12 days in culture. Transcript levels are normalized by *GAPDH* expression. The table shows relative gene expression, the percentage change (Δ%) and fold change in expression between the transduced and the control cultures, as well as the average cell VCN in bulk samples. Data represent single values from one experiment. E: PCR amplification efficiency; R^2: square of Pearson's correlation coefficient.

3.5.2.2 Erythroid liquid culture – Protocol B

In parallel to experiments performed with Protocol A, we tested the efficiency of Protocol B to achieve mass production of erythroid progenitors for transduction.

The primary cell culture system described by Fibach [438, 453, 454] is a two-phase culture method for the growth of a large and relatively pure population of erythroid cells. It has been successfully established in our laboratory for studying the effect of pharmacological agents on HbF expression. During Phase-I culture, PB mononuclear cells (PBMCs) separated by density-gradient centrifugation are grown in the presence of haematopoietic cytokines such as SCF and cyclosporine A but in the absence of EPO. SCF will stimulate proliferation of haematopoietic progenitors while cyclosporine A will suppress proliferation of lymphocytes. In Phase-II culture, the erythroid progenitors from the previous phase differentiate in the presence of EPO into orthochromatic normoblasts and enucleated erythrocytes. For the application of gene transfer procedures, the erythroid liquid culture protocol was modified as previously described [455], which, however, does not provide specifics of this procedure. The proposed scheme of the procedure is in **Figure 3.5.6** below.

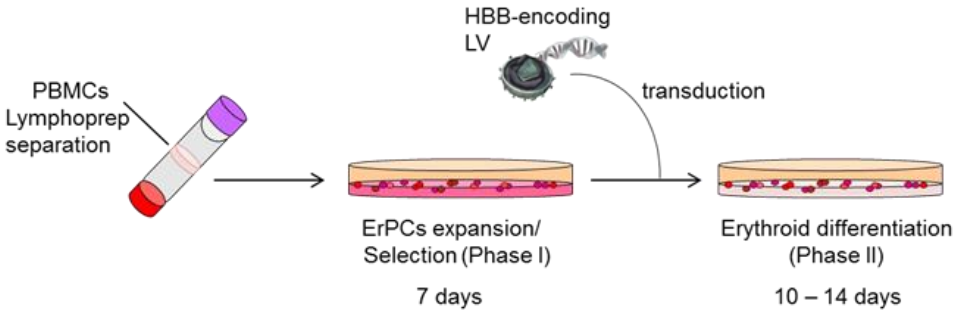
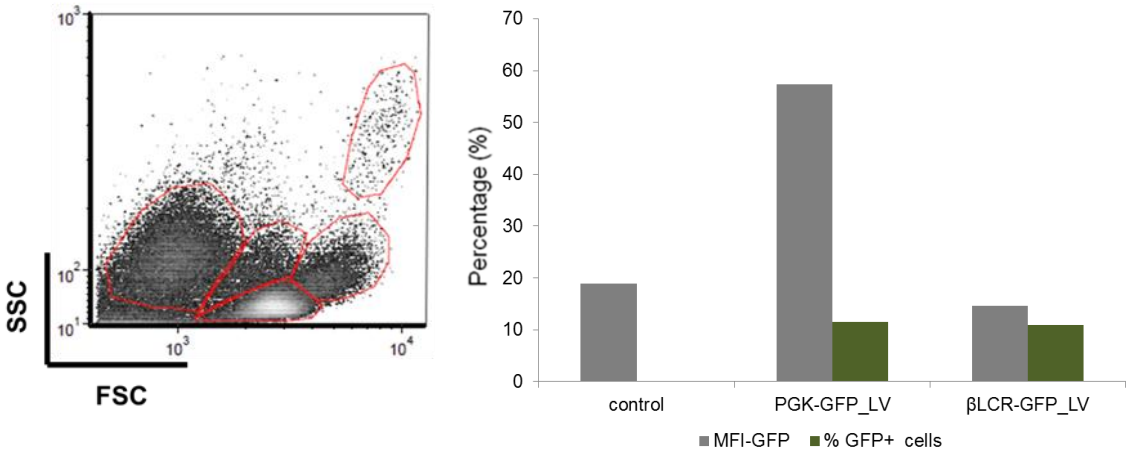


Figure 3.5.6. Protocol B. Outline of the cell culture procedure adapted from Breda et al [455]. PBMCs isolated on Lymphoprep density medium are maintained in EPO-independent Phase-I culture for 7 days. Thereafter, non-adherent cells (in majority CFU-Es) are collected and stimulated with cytokines for 24 h prior to transduction at various MOIs, with or without retronectin. Transduced cells are re-cultured in EPO-supplemented Phase-II medium for 12–14 days.

$CD34^+$ progenitor cells are the target for GT. Using this two-phase method, the majority of cells at the recommended transduction stage have already differentiated into late-erythroid CFU-E-

like progenitors, which are characterised by a $CD45^{+}GPA^{-}IL\text{-}3R^{-}$**$CD34^{-}$**$CD36^{+}CD71^{high}$ phenotype [456]. Ideally, pre-clinical investigation would target early erythroid-committed progenitors, BFU-E, which are characterized by a $CD45^{+}GPA^{-}IL\text{-}3R^{-}$**$CD34^{+}$**$CD36^{-}CD71^{low}$ phenotype [456], or even earlier progenitors. These cells are expanded during early stages of Phase-I culture but begin differentiation at an unspecified time point during Phase-I. In addition, the purity of the target cell population is a prerequisite for successful GT, and since BFU-Es are co-cultured with leukocytes and monocytes during Phase-I cultures, the resulting mixed cell population would make a poor target for transduction. Accordingly, we introduced a series of modifications to standard protocol procedures. First, we introduced positive magnetic isolation of $CD34^{+}$ cells from human PBMCs. $CD34^{+}$ cells ($0.5x10^{6}$) were pre-stimulated with a cytokine cocktail for 24 h and transduced on retronectin-coated plates using the MA821Qsh_B1 vector at a MOI of 100. The experiment also included a non-transduced sample. Both transduced and non-transduced cells represented the starting cell population for Phase-I culture. The modified protocol was set-up in parallel to the standard two-method culture in order to validate the inherent behaviour of the culture. Analyses showed that the two-phase culture supported the proliferation and differentiation of erythroid progenitors from a heterogeneous population of PBMCs, but failed to support viability with a starting sample of $CD34^{+}$-enriched cells for both, transduced and non-transduced, samples. As further protocol modifications would consume time and resources, we next investigated the success of the modified protocol described by Breda and colleagues, which incorporated transduction of cells at the beginning of Phase-II of the culture protocol [455]. At the end of Phase-I, the non-adherent cells in the supernatant, comprising mostly haematopoietic progenitors, were collected and transduced with a i) PGK-GFP vector for constitutive GFP expression and ii) βLCR-GFP vector for stage-specific erythroid specific GFP expression. Cells ($0.5x10^{6}$) were transduced at a MOI of 30 as described in **section 2.9.5.3.3**. The transduced and non-transduced cells ($0.125x10^{6}$/mL) were cultured in Phase-II medium. The level of cell death in the transduced samples (20.77±6.5%, n=2) was ~3.2 times that of the non-transduced sample (6.52%). However, we detected a substantial cell loss of >50% in both the transduced and the non-transduced samples, indicating that the process of transduction, rather than viral infection, was detrimental to cell viability. Also, cells were not selected for particular lineage-marker expression prior to transduction, raising the

possibility that cells other than erythroid-committed progenitors were exposed to the virus. The success of the transduction procedure was investigated via analysis of gene transfer using flow cytometry for GFP quantification (**Figure 3.5.7**). Constitutive GFP expression was ~3 times (+204%) the background level detected for the control, while stage-specific GFP expression was markedly below background levels (23%). While an unusually high level of autofluorescence might have contributed to these findings, it appeared from this experiment that transduction with the βLCR-GFP vector was either unsuccessful or that the transduced cells were not efficiently differentiating in Phase-II media. Surprisingly, both transductions resulted in ~11% of analysed cells being positive for GFP expression. Efficiency of the transduction process using low MOI (0.5–5) had previously been documented [180], but in the absence of sufficient cell material for qPCR-based VCN quantitation, low gene transfer efficiency for both vectors appeared the most likely explanation for the observed result. The flow cytometric analysis of transduced and non-transduced cells separated the population into 4–5 distinct clusters on an FSC vs. SSC plot (**Figure 3.5.7**). Phase-II culture exclusively supports the development of erythroid precursors, with myeloid cells reduced to <2% [454]. We could infer that these clusters represented morphologically distinct erythroid precursors rather than any other cell type, but had insufficient material for a confirmation by immunophenotyping or brightfield microscopy after cytocentrifugation. Overall, these data indicated the need for protocol optimization or troubleshooting of components for Protocol B, a daunting task in the absence of detailed protocol guidelines.



	MFI-GFP		% GFP⁺ cells
	MFI	Δ% (fold change)	
control	18.88		0
PGK-GFP_LV	57.39	203.97 (3.04)	11.49
βLCR-GFP_LV	14.58	-22.78 (0.77)	10.97

Figure 3.5.7. Transduction efficiency of erythroid progenitors using GFP-expressing LVs. PBMCs from normal donors were maintained in Phase-I medium for 7 days. Erythroid progenitors were harvested and stimulated with cytokines for 24 h prior to transduction with the PGK-GFP and βLCR-GFP LVs at a MOI of 30. Transduced cells were re-cultured in Phase-II medium and analysed by flow cytometry 9 days later. The histogram represents values of GFP mean fluorescence intensity (MFI-GFP) and the percentage (%) of GFP^+ cells. The table shows MFI-GFP, the percentage change (Δ%) and fold change in MFI-GFP between the transduced and the control cultures, and the percentage (%) of GFP^+ cells. Data represent single values from one experiment. The FSC vs. SSC dot plot separates the cell population into different discernible clusters, as indicated.

3.5.2.3 Erythroid liquid culture – Protocol C

The primary cell culture model introduced by Breda and colleagues is a modified version of the two-phase liquid culture system established by Fibach [454]. The key feature of this protocol is an initial $CD34^+$ cell expansion step that could relieve the bottleneck of having too few cells for transduction and subsequent analyses. The basic procedure involves the isolation of an enriched-$CD34^+$ cell population from PB (30 mL) and their expansion in StemSpan^{TM} expansion media supplemented with a cocktail of recombinant human cytokines (CC-100) and other additives formulated to selectively expand $CD34^+$ stem and progenitor cells. The proposed scheme of the procedure is shown in **Figure 3.5.8**.

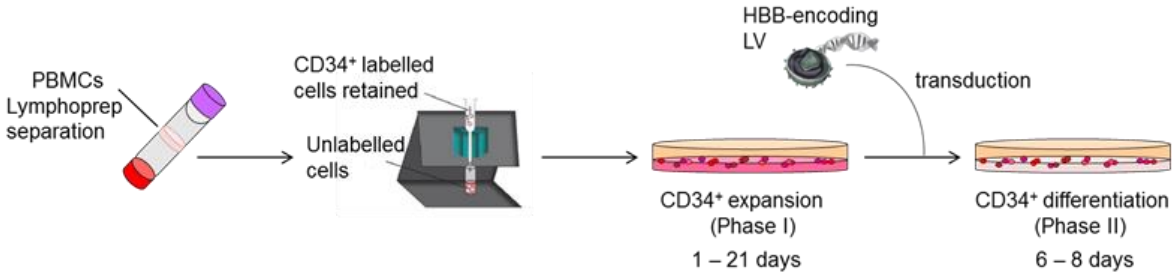


Figure 3.5.8. Protocol C. Outline of the cell culture procedure adapted from Breda et al [180]. $CD34^+$ cells are purified from the PBMC fraction isolated on Lymphoprep density medium by MACS affinity matrix selection. $CD34^+$ cells are expanded in Phase-I medium for up to 21 days, harvested and transduced at various MOIs. After 6 h of cell-virus hourly mixing, cells are re-cultured for erythroid differentiation in Phase-II medium for 6–8 days.

3.5.2.3.1 Protocol optimization using erythroid progenitors from normal donors

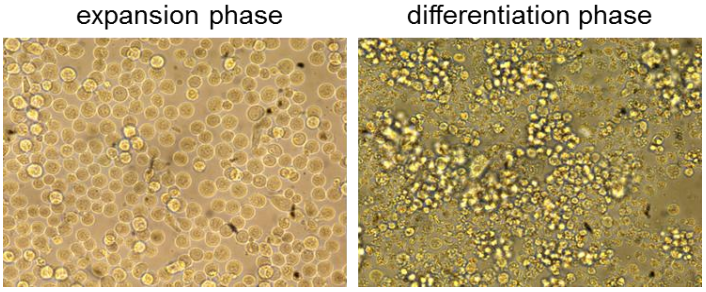
Initial cultures were based on normal $CD34^+$ progenitor cells in order to assess the effectiveness of the system to support cell proliferation and differentiation *in vitro*. We found that $CD34^+$-enriched cells isolated from moderate volumes (20–40 mL) of PB could result in a large number ($>5x10^6$) of highly homogeneous erythroid progenitor cell cultures after 8–10 days of expansion in Phase-I medium. Analysis of one representative experiment (**Figure 3.5.9**) showed that an inoculum of $0.2x10^6$ $CD34^+$ cells (collected on expansion day 7) rose to $2.9x10^6$ cells (+1440% expansion) at day 11 with minimal cell death (~6%). $CD34^+$ cells were harvested and re-cultured for an additional 4 days in Phase-II medium, during which they continued to proliferate, reaching

to 9.3x10^6 cells (4550% expansion) by day 15. The observed increase in total cell number was associated with a large accumulation of differentiated cells and a slight increase in cell death to 12%. Hb-containing cells were identified by benzidine staining and constituted ~97% of the culture. Cell morphology was analysed by May-Grünwald-Giemsa staining, indicating a high number of late orthochromatophilic erythroblasts (~67%) and fewer enucleated erythrocytes (~13%) while the remaining cells were in majority polychromatophilic normoblasts (~15%) and fewer proerythroblasts (~2.2%). At the end of the culture procedure, cell material was harvested for globin mRNA and protein analyses by RT-qPCR and RP-HPLC, respectively. Of note, high basal HBG mRNA levels were detected in cultures of adult erythroid progenitor cells, which was consistent with previous observations by others for this protocol [119, 180, 457]. Overall, preliminary experiments indicated that Protocol C produces sufficient cell material to perform all necessary biochemical and molecular analyses.

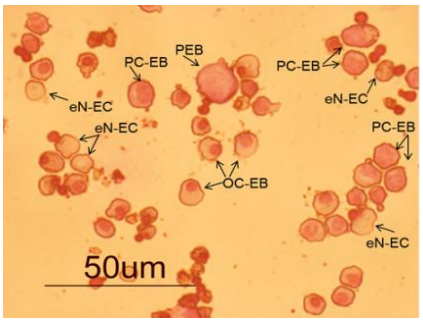
	blood (mL)	blood type	expansion (Days)	CD34$^+$ cell count
buffy coat	50		11	6.17 x10^6
	40	B$^+$	7	3.94 x10^7
peripheral blood	40	AB$^+$	10	7.60 x10^6
	20	O$^+$	9	8.24 x10^6

	CD34$^+$ cell count					
	expansion phase		differentiation phase			
	Day 7	Day 11	Day 12	Day 13	Day 14	Day 15
CD34$^+$ cell count (x10^6)	0.2	2.9	2.6	4.0	5.9	9.3
Δ%		1440.0	1200.0	1900.0	2850.0	4550.0
fold expansion		14.5	13.0	20.0	29.5	46.5
cell death(%)		6.0		6.2	6.8	11.8
haemoglobinized cells (%)						96.8

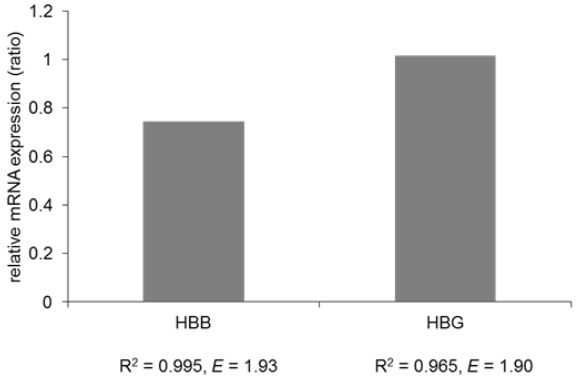
A



B



C



D

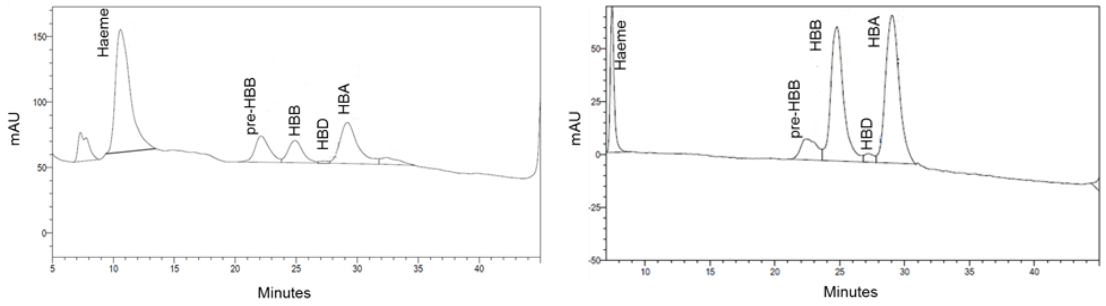


Figure 3.5.9. *In vitro* proliferation and erythroid differentiation of CD34^{+} cells. CD34^{+} cells isolated from normal donors were expanded in Phase-I medium for up to 12 days and analysed by trypan blue staining. Examples of cell counts are shown in the top table. An inoculum of 2x10^{5} cells was further expanded in fresh Phase-I medium and differentiated in Phase-II medium for 4 days prior to analysis. The bottom table shows the cell count on different days of the two-phase culture as determined by trypan blue staining. The percentage change (Δ%) and fold change in viable cell counts, as well as the percentage (%) of cell death are calculated. The percentage (%) of haemoglobinization is determined by benzidine staining. **A)** Optical examination of cultures by light microscopy. Phase-I culture (expansion) consists of a highly homogeneous cell population of big, round cells, while Phase-II culture (differentiation) is populated by dense aggregates of Hb-containing erythroblasts. **B)** Cell morphology assessment by May-Grünwald-Giemsa staining (magnification: x40; scale: 50 µm). **C)** Analysis for HBB and HBG mRNA transcript levels normalized to *GAPDH* by RT-qPCR. **D)** Analysis for HBA, HBB, HBD, and HBG1/HBG2 chains by RP-HPLC (left: normal primary erythroid culture; right: adult normal PB). proerythroblasts: PEB; basophilic erythroblasts: B-EB; polychromatophilic erythroblasts: PC-EB; orthochromatophilic erythroblasts: OC-EB; enuclated erythrocytes eN-EC; leukocytes: Leuk.; mAU: milli absorbance units; E: PCR amplification efficiency; R^2: square of Pearson's correlation coefficient.

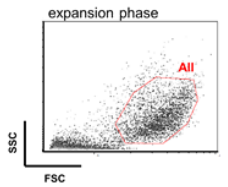
In order to determine the gene transfer efficiency, $CD34^+$ cells from normal PB were transduced with the βLCR-GFP-expressing vector at different MOIs and analysed for GFP expression (indicative of the efficiency of gene transfer) in mass cultures by flow cytometry (**Figure 3.5.10**). $CD34^+$ cells ($2.5x10^5$) were transduced at various vector concentrations (MOIs of 5, 10, 20, 30, 40, 50, and 100) to determine the amount of viral particles required to maximally transduce human progenitor cells. For average cell VCN analysis by qPCR, an inoculum of transduced $CD34^+$ cells ($0.5x10^5$) was cultured in expansion-phase medium for 3–4 days (**Figure 3.5.10A**) in order to avoid cell harvest after initiation of erythroid differentiation, which we had observed to compromise cell viability (data not shown). At the end of expansion phase, we noted an increase in total cell numbers by 538% for the non-transduced control ($3.19x10^5$) and by 306.4±102.7% for the transduced samples ($2.03±0.5x10^5$, n=7). In addition, trypan blue exclusion assays showed cell viability decreased by 7% in transduced cultures compared to control (73.6±8.8% vs. 81.1%, respectively). Expansion did not produce the minimum of $0.5x10^6$ cells required for DNA extraction using standard procedures, thus preventing reliable VCN determination. Flow cytometry analysis was conducted on portions of the expanded cells, showing efficient gene transfer for all transductions compared to the non-transduced control in the range of 22.5%–49.5% of GFP^+ cells and in positive correlation with increasing MOIs (logarithmic regression line: R^2=0.972). Of note, an MOI increase from 50 to 100 did not result in proportional increase in the percentage of GFP^+ cells. Expression from the βLCR-GFP vector is triggered most efficiently during late erythroid differentiation and to some extent modulated by the chromatin status of the proviral integration site and interaction by multiple (often random) integrations, which may contribute to non-detection of a proportion of transduced cells by flow cytometry. The remaining transduced $CD34^+$ cells ($2.0x10^5$) were cultured in differentiation-phase medium and harvested on day 10 of Phase-II for end-point analyses (**Figure 3.5.10B)**. We noted a low number of viable cells, with 74.4±5.4% cell loss for transduced cultures ($0.51±0.11x10^5$ live cells with 46.2±4.2% viability, n=7) and 64% cell loss for the non-transduced control ($0.72x10^5$ live cells with 55.4% viability). Given the scarcity of the cell material, samples were only used to quantify the percentage of GFP^+ cells and MFI-GFP by flow cytometry. Flow cytometry analysis separated the cell population into two clusters on the FSC vs. SSC plot

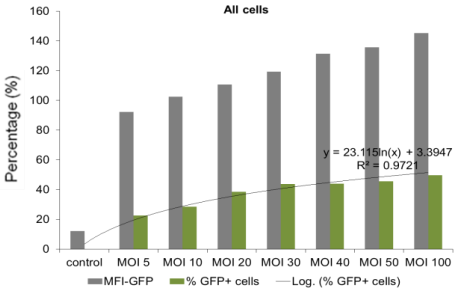
(**Figure 3.5.10B**). The SSC^{high}/FSC^{low} sub-population constituted 53%–68% of the total population and exhibited a relatively low gene transfer efficiency across the different MOIs (4%–10% GFP^+ cells; 31.08±1.65 MFI-GFP, n=7). Conversely, the SSC^{low}/FSC^{high} sub-population constituted 25%–35% of the total population and exhibited a higher gene transfer efficiency across the different MOIs (38.5%–73.6% GFP^+ cells; 282.1±43.3 MFI-GFP, n=7). We noted an increasing proportion of GFP^+ cells with increasing MOIs (logarithmic regression line: R^2=0.952) and a higher, albeit non-proportional, MFI-GFP with increasing MOIs (logarithmic regression line: R^2=0.8862). A SSC^{low} feature is indicative of cells with reduced structural complexity, as would be the case with differentiating cells heading towards enucleation and endoplasmic reticulum loss during erythroid maturation. As GFP expression was under the erythroid-specific βLCR regulatory element, only those cells driven towards terminal erythroid differentiation would strongly fluoresce. Therefore it is possible that the SSC^{low} cluster represented differentiated cells.

We presented correlation using the logarithmic trendline function of *Excel* since the data would exhibit a near-linear change at first (i.e.: GFP expression level increasing proportionally to the vector input) and would then asymptotically approach a saturation threshold of the transduction efficiency. Only a small difference (~2.6%) in GFP^+ cells was detected between a MOI of 50 and 100, indicating that a MOI of 50 already approached maximal transduction, with a further increase in MOI giving diminishing returns. At higher MOI, an increasing number of cell have multiple integrations [458], with a minor effect on βLCR-derived expression levels but an increasing likelihood of insertional mutagenesis.

A

	Day 0	expansion phase _ Day 4							
		control	βLCR-GFP_LV						
			MOI 5	MOI 10	MOI 20	MOI 30	MOI 40	MOI 50	MOI 100
CD34⁺ cell count (x10⁵)	0.50	3.19	2.26	1.55	1.15	2.10	2.60	2.46	2.12
% change		538	352	210	130	320	420	392	324
fold expansion		7.87	6.24	5.69	3.00	5.32	6.73	6.32	5.43
viability (%)		81.10	72.44	54.31	76.64	78.95	77.12	77.85	77.97

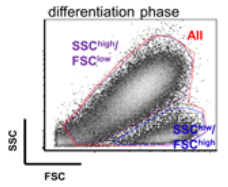




	MFI-GFP		% GFP⁺ cells
	MFI	Δ% (fold change)	
control	12.07		0
MOI 5	92.16	663.55 (7.64)	22.57
MOI 10	102.53	749.46 (8.49)	28.54
MOI 20	110.7	817.15 (9.17)	38.48
MOI 30	119.32	888.57 (9.89)	43.69
MOI 40	131.31	987.90 (10.88)	43.98
MOI 50	135.59	1023.36 (11.23)	45.44
MOI 100	145.17	1102.73 (12.03)	49.58

B

	Day 0	differentiation phase _ Day 10						
		control	βLCR-GFP_LV					
			MOI 5	MOI 10	MOI 20	MOI 30	MOI 40	MOI 50	MOI 100
CD34⁺ cell count (x10⁵)	2.00	0.72	0.68	0.56	0.40	0.58	0.36	0.50	0.50
% change		-64	-66	-72	-80	-71	-82	-75	-75
fold expansion		0.36	0.34	0.28	0.20	0.29	0.18	0.25	0.25
viability (%)		55.38	52.31	51.85	41.67	45.31	45.00	43.10	43.86



	All cells			SSC^high/FSC^low			SSC^low/FSC^high		
	MFI-GFP		% GFP⁺ cells	MFI-GFP		% GFP⁺ cells	MFI-GFP		% GFP⁺ cells
	MFI	Δ% (fold change)		MFI	Δ% (fold change)		MFI	Δ% (fold change)	
control	15.75		0	16.85		0.00	10.2		0
MOI 5	56.69	455.78 (5.56)	7.88	29.31	187.35 (2.87)	3.44	221.62	2072.75 (21.73)	38.51
MOI 10	56.43	453.24 (5.53)	12.07	30.73	201.27 (3.01)	6.65	230.84	2163.14 (22.63)	48.18
MOI 20	52.33	413.04 (5.13)	11.37	29.62	190.39 (2.90)	6.14	280.24	2647.45 (27.47)	60.74
MOI 30	56.67	455.59 (5.56)	12.03	31.28	206.67 (3.07)	6.63	290.66	2749.61 (28.50)	65.28
MOI 40	54.67	435.98 (5.36)	14.2	30.07	194.80 (2.95)	8.40	304.51	2885.39 (29.85)	66.69
MOI 50	56.98	458.63 (5.59)	15.58	33.67	230.10 (3.30)	9.77	301.19	2852.84 (29.53)	70.91
MOI 100	52.54	415.10 (5.15)	15.79	32.86	222.16 (3.22)	10.53	345.63	3288.53 (33.89)	73.57

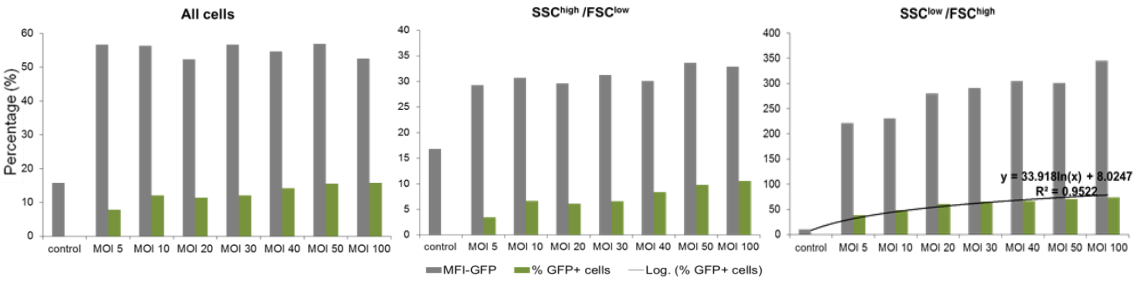


Figure 3.5.10. Gene transfer efficiency in CD34⁺ cells using the βLCR-GFP LV. CD34⁺ cells were transduced with the βLCR-GFP LV at various MOIs, as indicated, and cultured in **A)** Phase-I medium for 4 days and **B)** Phase-II medium for 10 days, and analysed for cell counts by trypan blue staining, as well as GFP mean fluorescence intensity (MFI-GFP) and the percentage (%) of GFP⁺ cells by flow cytometry. The tables show MFI-GFP, the percentage change (Δ%) and fold change in MFI-GFP between the transduced and the control cultures, for the total cell population (All cells) and each sub-population (SSC/FSC) separately. Data are presented for single transductions from one experiment. The square of Pearson's correlation coefficient (R^2) is calculated using the logarithmic trendline function of *Excel*.

To avoid dependence of flow-cytometric VCN estimates on erythroid differentiation, we chose the PGK-GFP-expressing vector for subsequent experiments, which uses the constitutive PGK promoter for ubiquitous GFP expression. For an additional comparison of MOI 50 and 100, we assumed that the number of transduced cells ($0.25x10^5$) in the previous experiment had been too small to achieve cell expansion in numbers sufficient to perform average cell VCN quantification, proceeding for this additional measurement with transduction of $1x10^6$ cells (**Figure 3.5.11**). An inoculum of transduced $CD34^+$ cells ($0.2x10^6$) was cultured in expansion-phase medium for 5 days, resulting in 280%–385% higher cell numbers, which were sufficient to perform VCN quantification (**Figure 3.5.11A**). More specifically, transduction using MOI of 100 ($0.76x10^6$; 280%) resulted in half the cell expansion observed for non-transduced control ($1.35x10^6$; 575%) and exhibited comparable viability at approximately 73.8±3.6%. Transduction using an MOI of 50 ($0.97x10^6$; 385%) resulted in 33% lower cell expansion compared to the non-transduced control and in 37.5% higher cell expansion compared to transduction using an MOI of 100. Also, using an MOI of 50, the transduced cells exhibited ~14% greater viability (84.1%) than for the other two conditions indicating that cell loss may have occurred as a result of metabolic acidosis caused by active cell growth or by multiple transductions at high LV-cell ratio. We measured an average cell VCN of 44.4 and 66.5 in bulk cultures of cells transduced with an MOI of 50 and 100, respectively. Gene transfer was efficient for both transductions, resulting in 43.34% (MOI of 50) and 60.21% (MOI of 100) of GFP^+ cells. The experiment showed an expected increase of GFP^+ cells (38.9%) with increasing VCN (49.6%) from an MOI of 50 to 100 (logarithmic regression line: R^2=0.9903) and a corresponding increase in MFI-GFP, between transduced and non-transduced cells, from 651.1% (MOI 50) to 795.7% (MOI 100). Since the increase in MFI-GFP (19.2%) was not proportional to increasing VCN (49.6%) from an MOI of 50 to 100 (logarithmic regression line: R^2=0.9597), this suggested that a large number of GFP^+ cells would bear multiple integrations for MOI 100.

Measurements by flow cytometry were complemented by qPCR, which is independent of vector expression but only allows analysis of the cell population in bulk (**Figure 3.5.11A**) [459]. Under these conditions (transduction and culture protocols) the PGK-GFP vector using an MOI of 100 resulted in ~60% of GFP^+ cells. Increasing the MOIs above 100 could have elevated gene

transfer efficiency, but also likely cytotoxicity. In this experiment we once again observed saturation between MOI 50 and 100 for both, VCN and MFI-GFP (i.e. a disproportionately lower increase for both parameters for a doubling of viral load), in line with observations for the βLCR-GFP vector.

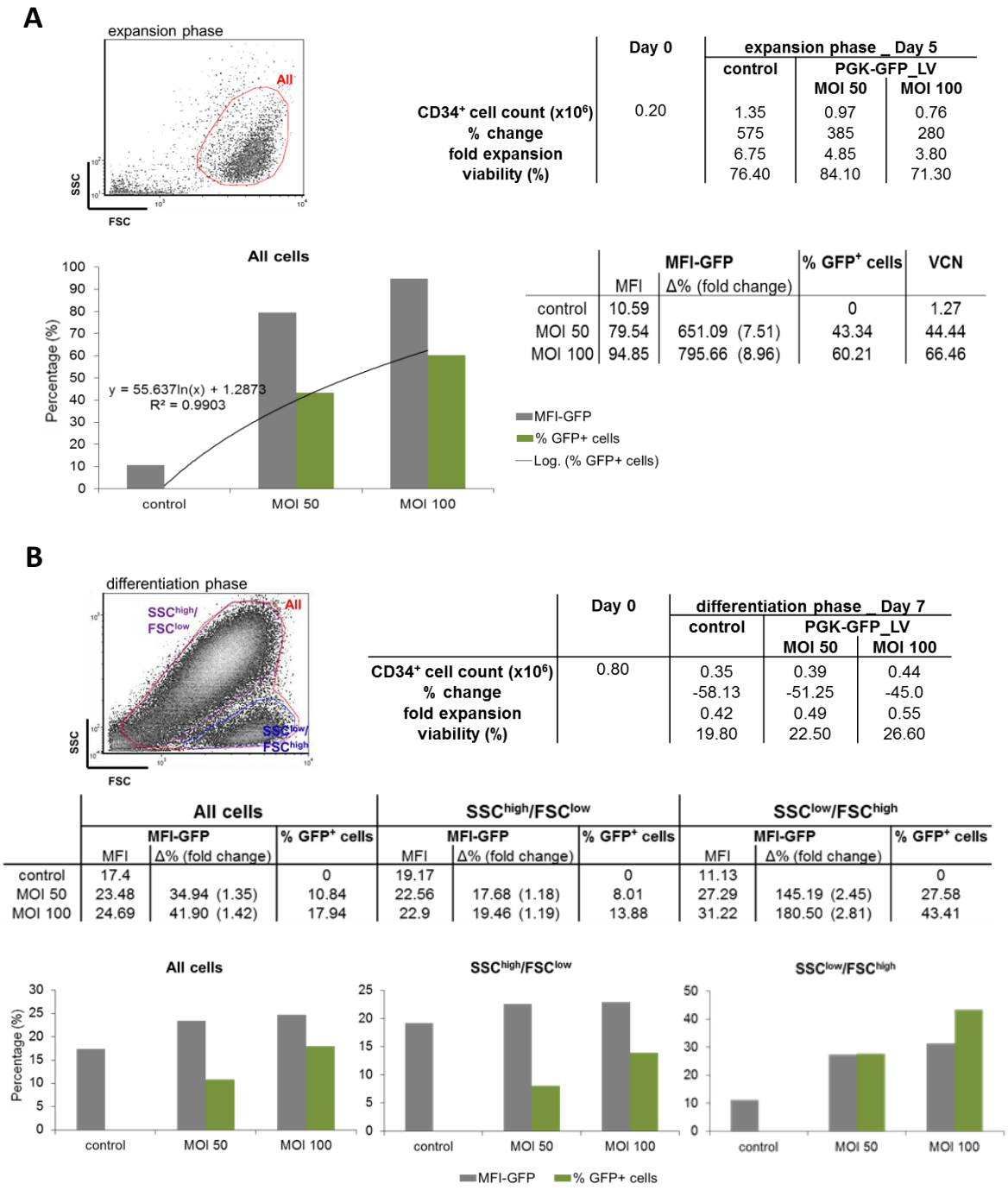
The remaining transduced $CD34^+$ cells ($0.8x10^6$) were cultured in differentiation-phase medium and harvested on day 7 for end-point analyses (**Figure 3.5.11B**). We noted a low number of viable cells, which was calculated as a 51.5±6.6% loss across the control and transduced cultures. Given the resulting scarcity of the cell material ($3.88±0.5x10^5$, n=3), analysis was performed only for GFP expression using flow cytometry. Flow cytometry analysis separated the cell population into two clusters on the FSC vs. SSC plot, and revealed a higher MFI-GFP for the SSC^{low}/FSC^{high} cluster. This cluster constituted ~30% of the total cell population and exhibited a higher percentage of GFP^+ cells for both MOI 50 (27.6%) and MOI 100 (43.4%), which was, nonetheless, smaller than that observed for undifferentiated cells (see previous paragraph). The large proportion of dead cells in (22.9±3.4% viability, n=3), however, may have masked the true level of GFP expression, with transduced cells losing (most of) their GFP fluorescene upon apoptosis or necrosis.

Lastly, we noted that the SSC^{low}/FSC^{high} clusters in both studies (i.e. for both, the βLCR-GFP and the PGK-GFP vector) exhibited high transduction efficiency (see scatter plots in **Figures 3.5.10 and 3.5.11**). Since Phase-I medium is formulated to specifically support growth of erythroid-committed progenitors and, consequently, to contribute to a highly homogenous population of cells at the time of transduction, we rejected the possibility of a bias in transducibility or promoter activity for a particular starting cell type. Conversely and assuming that differences in SSC indicate different stages of differentiation, these observations might instead represent differential vector-derived expression, brought about by changes in the chromatin structure and promoter activity during differentiation. One of the main concerns of this study was the low rate of expansion and the high rate of cell death during differentiation. Thus more $CD34^+$ cell cultures were established but with limited success (experiments not shown). As a result, I visited Dr. Stefano Rivella's laboratory (Weill Cornell Medical College, New York), for a 6-week training programme on the expansion and differentiation of stem cells in the two-

phase liquid culture protocol. The system was subsequently successfully established under my training and tested with substantial contribution by other laboratory members to function efficiently and reproducibly in our group at CING (**Table 3.5.3**).

sample	sex	genotype		β-thalassaemia type	expanded $CD34^+$ cells ($x10^6$)
	(M/F)	α-thalassaemia	β-thalassaemia		
1	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	24.2
2	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	14.5
3	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	9.5
4	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	27.7
5	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	9.5
6	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	17.3
7	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	30
8	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	24
9	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	28.9
10	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	12
11	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	14.2
12	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	6.4
13	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	13.5
14	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	28.5
15	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	23
16	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	25.4
17	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	24.1
18	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	33.6
19	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	22.8
20	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	20.3
21	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	7.6
22	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	11.3
23	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	38.4
24	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	30
25	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	35
26	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	10
27	M	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	28.4
28	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	25
29	F	αα/αα	IVSI-6 / IVSI-110	β^+/β^+	26.6
30	F	αα/αα	N/N	Normal	8.7

Table 3.5.3. Thalassaemia samples. Erythroid progenitor cells isolated on Lymphoprep density medium were expanded *in vitro* using Protocol C. The cell number is indicative for Phase-I expansion days 8–10. N: normal; IVSI-110: HBB:c.93-21G>A; IVSI-6: HBBc.92+6T>C.



	Day 0	expansion phase _ Day 5		
		control	PGK-GFP_LV	
			MOI 50	MOI 100
CD34⁺ cell count (x10⁶)	0.20	1.35	0.97	0.76
% change		575	385	280
fold expansion		6.75	4.85	3.80
viability (%)		76.40	84.10	71.30

	MFI-GFP		% GFP⁺ cells	VCN
	MFI	Δ% (fold change)		
control	10.59		0	1.27
MOI 50	79.54	651.09 (7.51)	43.34	44.44
MOI 100	94.85	795.66 (8.96)	60.21	66.46

	Day 0	differentiation phase _ Day 7		
		control	PGK-GFP_LV	
			MOI 50	MOI 100
CD34⁺ cell count (x10⁶)	0.80	0.35	0.39	0.44
% change		-58.13	-51.25	-45.0
fold expansion		0.42	0.49	0.55
viability (%)		19.80	22.50	26.60

	All cells			SSC^{high}/FSC^{low}			SSC^{low}/FSC^{high}		
	MFI-GFP		% GFP⁺ cells	MFI-GFP		% GFP⁺ cells	MFI-GFP		% GFP⁺ cells
	MFI	Δ% (fold change)		MFI	Δ% (fold change)		MFI	Δ% (fold change)	
control	17.4		0	19.17		0	11.13		0
MOI 50	23.48	34.94 (1.35)	10.84	22.56	17.68 (1.18)	8.01	27.29	145.19 (2.45)	27.58
MOI 100	24.69	41.90 (1.42)	17.94	22.9	19.46 (1.19)	13.88	31.22	180.50 (2.81)	43.41

Figure 3.5.11. Gene transfer efficiency in CD34$^+$ cells using the PGK-GFP LV. CD34$^+$ cells were transduced with the PGK-GFP vector at a MOI of 50 and 100 as indicated, and cultured in **A)** Phase-I medium for 5 days and **B)** Phase-II medium for 7 days, and analysed for cell count by trypan blue staining, as well as GFP mean fluorescence intensity (MFI-GFP) and the percentage (%) of GFP$^+$ cells by flow cytometry. The tables show MFI-GFP, the percentage change (Δ%) and fold change in MFI-GFP between the transduced and the control cultures, for the total cell population (All cells) and each sub-population (SSC/FSC) separately, as well as the average cell VCN in bulk samples. Data are presented for single transductions from one experiment. The square of Pearson's correlation coefficient (R^2) is calculated using the logarithmic trendline function of *Excel*.

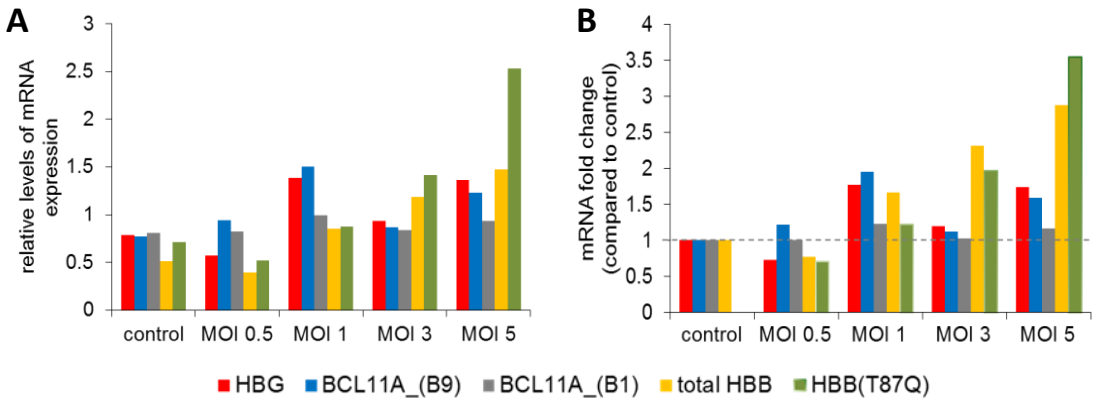
3.5.2.3.2 Transduction using GT vectors of erythroid progenitor cells from β-thalassaemia carriers

Experiments were performed with $CD34^+$ progenitor cells isolated from a person with β-thalassaemia trait (β^+/β^N) or thalassaemia major (β^+/β^0). At the end of the standard culture procedure, total HBB (endogenous and vector-encoded), HBG and BCL11A mRNA content was assayed by RT-qPCR. When sufficient cell material was produced, cell lysates were prepared for protein analysis using Western blotting. The first experiment involved single-vector transductions at different MOI in order to determine the amount of viral particles required to maximally transduce human progenitor cells using Protocol C. Thereafter, a total of five experiments were set up, two of which failed because of inadequate erythroid differentiation with <40% of cells scoring positive for benzidine Hb staining and/or massive cell death (data not shown). Data from the remaining three experiments (>90% haemoglobinization) were pooled and subjected to outlier detection and removal in order to reduce variance and enhance the power of statistical analysis. Data were analysed using one-way ANOVA.

Initially, $CD34^+$ cells from a person with β-thalassaemia trait (IVSI-110/N) were transduced with the MA821Qmir_B9 vector at various concentrations (MOIs 0.5, 1, 3, and 5) (**Figure 3.5.12**), to determine the optimal conditions for producing cells with a VCN between 0.1 and 2, which is the expected range of transduction in a clinical setting [180, 240]. In contrast to the previous transduction protocol, which adds virus of an MOI 100 on retronectin-bound cells, Protocol C uses a low MOI of vectors with $>10^8$ TU/mL and a 6-h incubation of the cell-virus mix with hourly resuspension for increased transduction efficiency (Laura Breda, personal communication). The total level of HBB mRNA, which takes into account both endogenous and vector-encoded HBB mRNA, showed an increasing trend with increasing MOIs (R^2=0.902). Analysis of vector-encoded HBB using the T87Qx2x3 primer pair showed a similar increase in mRNA expression with increasing MOIs (R^2=0.783), as would be expected with higher transduction rates. Using an MOI of 0.5, we did not detect deviation of *HBB* expression from the basal levels in control $CD34^+$ cells. On the contrary, using an MOI of 5, vector-derived HBB mRNA made up 135.9% of total HBB mRNA (34.6% for MOI 1 and 74.7% for MOI 3), representing an exaggerated proportion of vector-derived transcripts assuming a VCN of 1. As a reference, May and

colleagues [182] reported that 1 VC/cell of TNS9 vector would express 16.8±3.9% of vector-encoded human HBB mRNA measured relative to the level of mouse Hbb mRNA in the spleen cells of BM chimeras transduced with TNS9. Romero and colleagues [460] reported that 1 VC/cell of CCL-βAS3-FB vector would express 26.22±10.71% of vector-encoded anti-sickling HBB mRNA in erythroid progenitor cultures of SCD donor BM-CD34^{+} cells. Arguably, cells with single vector integrants generally and also in my case express lower levels of vector-encoded HBB mRNA compared to endogenous HBB mRNA, since all LV-based *HBB* vectors, including MA821Q, express the *HBB* transgene under a reduced βLCR element. This prompted us to disqualify both, an MOI of 0.5 and 5, for subsequent experiments. According to [458], the provirus integration is more likely to occur during mitosis such that at a low MOI only one of the two daughter cells would carry a vector copy resulting in a theoretical maximal gene transfer of 50% with an MOI of 1. However, assuming a random distribution of viral particles to target cells, the use of 1 VCN per cell will in fact result in a still lower percentage of transduced cells (estimated 30%–37%, with 1.4–1.7 VC/cell for the transduced sub-population [458]). As a result, with 1 vector copy per cell in bulk samples we would not have expected to detect doubling of HBB output which was, nevertheless, observed using an MOI of 3. Therefore, we speculated that optimal transduction efficiency could be achieved between an MOI of 1 and 3, and thus we chose an MOI of 2 for subsequent experiments.

Lastly, we detected increased *HBG* expression in samples transduced at an MOI ≥1, even though there was no corresponding (and, according to the rationale for our shRNA expression strategy, “causative”) knockdown in BCL11A expression.



	relative gene expression										VCN
	HBG		BCL11A_(B9)		BCL11A_(B1)		total HBB		HBB^{T87Q}		
	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	
control	0.78		0.77		0.81		0.51		0.71		0.009
MOI 0.5	0.57	-26.96 (0.73)	0.94	21.91 (1.22)	0.82	1.73 (1.02)	0.40	-22.45 (0.78)	0.52	-27.51 (0.72)	0.55
MOI 1	1.39	77.23 (1.77)	1.51	94.88 (1.95)	0.99	23.29 (1.23)	0.85	66.02 (1.66)	0.88	22.84 (1.23)	0.52
MOI 3	0.93	19.48 (1.19)	0.86	11.86 (1.12)	0.84	3.84 (1.04)	1.18	131.45 (2.31)	1.42	98.25 (1.98)	3.76
MOI 5	1.36	73.96 (1.74)	1.23	58.91 (1.59)	0.94	16.20 (1.16)	1.47	187.43 (2.87)	2.53	254.76 (3.55)	5.25

Figure 3.5.12. RT-qPCR analysis of mRNA expression in transduced primary erythroid cultures. $CD34^+$ cells (IVSI-110/N) were transduced with the MA821Qmir_B9 LV at various MOIs as indicated, and analysed for HBG, BCL11A (regions surrounding the B9 and B1 recognition sites), total HBB (endogenous and vector-derived), and vector-derived HBB (HBB^{T87Q}) mRNA levels by RT-qPCR after 6 days in Phase-II medium. Values are normalized by *GAPDH* expression. The histograms are single values from one experiment and show **A)** raw data of expression and **B)** fold change relative to the control sample, which is arbitrarily set at a value of 1. The table also shows the percentage change (Δ%) in expression between the transduced samples and the control, as well as the average cell VCN in bulk cell samples.

In subsequent experiments (n=3), GT vectors were characterized in $CD34^+$ blood cells of IVSI-110/N and IVSI-110/IVSI-110 donors to address six experimental aims as shown in **Table 3.5.1**.

		GT and test vectors									
		MA821 ('GLOBE')	MA821Q	MA821Qsh_B9	MA821Qsh_B1	MA821QmiR_B9	MA821QmiR_B1	MA821QmiR_Scr	pLKOsh_B9	pLKOsh_B1	pLKOsh_Scr
experimental aims	vector-derived HBB^{T87Q} vs. base vector	x	x	x	x	x	x				
	effect of shRNA expression on vector-derived *HBB* levels		x	x	x	x	x				
	competition between vector-derived *HBB* and endogenous *HBG* genes	x	x	x	x	x	x		x	x	
	efficiency of U6-driven and intron-derived shRNA production, indirectly via the level of target mRNA and reactivation of *HBG* expression		x	x	x	x	x	x	x	x	x
	silencing efficiency of conventional shRNA and miR30 target site flanked shRNA		x	x	x	x	x	x			
	non-specific silencing of BCL11A by control shRNAs							x			x

Table 3.5.1. The table shows the vectors tested in human erythroid progenitors to address a set of experimental aims. The correspondence of vector and aim is indicated by 'x'.

Relative expression values were determined as a ratio of target gene and reference gene (i.e., *GAPDH*) mRNA levels in order to correct for intra-assay variability. These values represent the raw data collected during RT-qPCR and are shown in **Figure 3.5.13** for each experiment separately. The quantity of the target gene expression in the transduced cell samples was first determined using a standard curve. Because the standard curve relies on a narrow range of dilutions of a pool of sample cDNAs, it does not represent a solid (non-changing) and reliable basis in order to compare expression data between different experiments, assuming differences in PCR amplification efficiency [461]. Therefore, we expressed the quantity of the target gene expression in the transduced samples relative to the control sample, which is arbitrarily set at a value of 1, such that all experimentally derived quantities were reported as a fold difference (expression ratio) between test samples and control. Thus a fold change >1 would mean that the gene is up-regulated, whilst a fold change <1 implies that the gene is down-regulated. Single observations of differential expression do not have any statistical significance. Therefore, in order to enable statistical analysis, we pooled the 'fold change' values for each experimental condition from the three different experiments, followed by outlier detection and removal, so that each experimental condition was represented by 2–3 biological samples for subsequent analyses. The percentage change in expression ratio was calculated for each LV-treated sample in relation to basal expression in the non-transduced control cells (set at a value of 1.0). Statistical significance was tested using one-way ANOVA (**Figure 3.5.14**).

We found a non-significant change in the HBB mRNA production in LV-treated and control cultures ($F(12,17)=2.088$, $p=0.0806$) (**Figure 3.5.14A**). This result showed that the difference in HBB mRNA between vector-encoded normal HBB and modified HBB derivative β^{T87Q} was mainly due to chance. Similarly, results suggested an insignificant impact of concomitant shRNA expression on vector-derived HBB levels. The expression level of vector-encoded *HBB* was deduced from the relative increase in the total HBB output. We had previously established an RT-qPCR assay using the T87Qx2x3 primer pair for specific quantification of the vector-encoded HBB. In this study, an amplification signal above the threshold of detection was produced only for a minority of samples with higher VCN, such that further analysis was not performed. Examples of pair-wise comparisons which resulted in large changes in *HBB*

expression are, e.g., the MA821Q vector (0.71±0.15) gave a 45% lower expression ratio than the MA821 vector (1.29±0.61). Also, the MA821QmiR_B9 (0.39±0.53) and the MA821Qsh_B9 (1.03±0.13) vectors gave rise to 45.6% lower and a 45% higher expression ratio, respectively, than the MA821Q vector. The lack of statistical significance compared to controls (with the above *p*-value of 0.0806), despite a general trend of increased *HBB* expression in HBB-transduced samples could be attributed to variability in expression values with only 2–3 samples included per experimental sample, as well as to an already high level of basal *HBB* expression among the mildly thalassaemic or carrier subjects.

Similarly, we obtained a non-significant change in HBG mRNA production of LV-treated and control cultures ($F(12,16)=0.403$, $p=0.9414$) (**Figure 3.5.14B**). Analysis revealed a strikingly lower *HBG* expression (by 52%) in samples transduced with the pLKOsh_B9 vector designed for BCL11A knock-down (0.47±0.67). The siRNA sequence included in the pLKOsh_B9 vector was previously shown to up-regulate *HBG* expression via BCL11A knockdown [119]. Additionally, the *HBG* expression ratio detected in samples transduced with the other BCL11A-specific shRNA or miRNA vectors (0.94±0.03–1.43±0.10) did not deviate significantly from the range of expression ratios detected in samples transduced with the control vectors (0.78±0.14–1.31±0.73).

We, next, investigated the silencing efficiency of conventional shRNA and miRNA hairpin designs via quantification of BCL11A mRNA expression. First, we analysed the data produced with the shB9 primer pair and found a statistically significant change in the BCL11A mRNA output ($F(12,16)=4.255$, $p=0.0041$) (**Figure 3.5.14C**). A Tukey's pairwise comparison test gave a statistically significant difference at $p<0.05$ for i) samples transduced with the pLKOsh_B9 (0.15±0.21) vs. the MA821 (1.48±0.76, -89.8%) and the MA821Q (1.65±0.16, -90.8%) vectors, ii) the MA821QmiR_B9 (0.23±0.32) vs. the MA821 (1.48±0.76, -84.1%) and the MA821Q (1.65±0.16, -85.8%) vectors, and iii) the co-transduced sample of MA821Q&pLKOsh_B1 (0.18±0.14) vs. the MA821 (1.48±0.76, -87.5%) and the MA821Q (1.65±0.16, -88.8%) vectors. In comparison to the basal expression in the control sample (arbitrarily set at a value of 1.0), the pLKOsh_B9 (0.15±0.21) and the MA821QmiR_B9 (0.23±0.32) vectors resulted in an 84.9% and 76.6% lower *BCL11A* expression, respectively, albeit without statistical significance due to the

large variation among samples. The MA821Qsh_B9 vector (0.81±0.22) gave a 19% lower *BCL11A* level of expression compared to control. Strikingly, the control hairpin vector pLKOsh_Scr (0.42±0.13) reported a larger effect of a 57.9% lower expression, while the MA821QmiR_Scr vector (1.06±0.13), which encodes an identical siRNA scramble sequence to the pLKOsh_Scr vector, did not reveal similar non-specific silencing. Also, the significantly lower *BCL11A* expression in the co-transduced MA821Q&pLKOsh_B1 sample was not detected in the sample transduced solely with the pLKOsh_B1 vector (0.69±0.18, -30.8%), at least using the shB9 primer pair. Therefore, we investigated *BCL11A* expression using the shB1 primer pair. Analyses revealed a non-significant increase in the BCL11A output ($F(12,14)=1.459$, $p=0.2475$) (**Figure 3.5.14D**). Of note, none of the vectors encoding the B1 hairpin (1.22±0.15–1.55±0.15) resulted in lower expression ratios than basal expression in the non-transduced control cells. Also, the control vectors pLKOsh_Scr (0.93±0.02) and MA821QmiR_Scr (0.97±0.05) resulted in a comparable 6.8% and 2.5% lower expression, respectively.

A

	relative gene expression								VCN
	HBG		BCL11A_(B9)		BCL11A_(B1)		total HBB		
	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	
control	0.17		0.95		0.80		0.37		0.06
MA821	0.10	-44.35 (0.56)	0.66	-30.98 (0.69)	0.89	10.74 (1.11)	0.26	-29.93 (0.70)	56.46
MA821Q	0.12	-30.48 (0.70)	0.75	-20.97 (0.79)	0.81	0.65 (1.01)	0.22	-39.71 (0.60)	67.61
MA821QmiR-Scr	0.08	-52.94 (0.47)	0.09	-90.46 (0.10)	0.81	1.33 (1.01)	0.11	-70.31 (0.30)	28.59
pLKOsh_Scr	0.15	-11.56 (0.88)	0.32	-67.00 (0.33)	0.74	-8.13 (0.92)	0.32	-11.38 (0.89)	0.88
MA821Qsh_B9	0.19	10.19 (1.10)	0.62	-34.57 (0.65)	0.85	5.47 (1.05)	0.33	-9.56 (0.90)	97.80
MA821QmiR_B9	0.0008	-99.54 (0.00)	0.005	-99.49 (0.01)	0.78	-3.22 (0.97)	0.004	-99.00 (0.01)	136.37
pLKOsh_B9	$3.14x10^{-06}$	-100.00 (0.00)	0.003	-99.67 (0.00)	0.96	19.53 (1.20)	0.003	-99.30 (0.01)	2.32
MA821Q + pLKOsh_B9	0.18	0.84 (1.01)	0.019	-98.05 (0.02)	0.81	1.05 (1.01)	0.07	-81.80 (0.18)	62.08

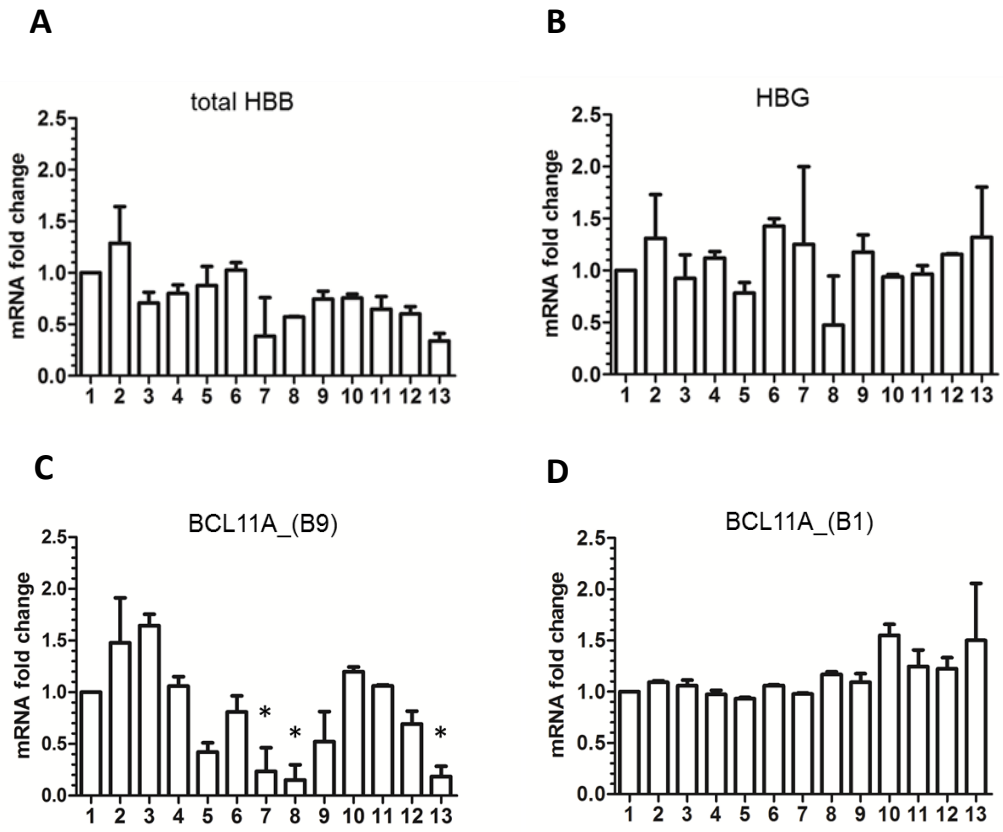
B

	relative gene expression								VCN
	HBG		BCL11A_(B9)		BCL11A_(B1)		total HBB		
	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	
control	0.79		0.81		0.76		1.11		0.06
MA821	1.07	35.79 (1.36)	1.24	54.19 (1.54)	0.90	18.71 (1.19)	1.37	23.53 (1.24)	0.18
MA821Q	0.90	15.11 (1.15)	1.24	53.59 (1.54)	1.03	36.25 (1.36)	0.90	-18.79 (0.81)	0.39
MA821QmiR-Scr	0.93	18.31 (1.18)	0.93	15.04 (1.15)	1.02	34.40 (1.34)	0.79	-28.33 (0.72)	0.25
pLKOsh_Scr	1.12	43.05 (1.43)	0.97	20.48 (1.20)	0.87	14.10 (1.14)	1.32	19.20 (1.19)	1.24
MA821Qsh_B9	1.06	35.33 (1.35)	0.78	-3.43 (0.97)	1.16	52.98 (1.53)	1.13	1.78 (1.02)	0.05
MA821QmiR_B9	0.91	15.90 (1.16)	0.37	-53.65 (0.46)	1.43	88.55 (1.89)	0.84	-23.97 (0.76)	8.39
pLKOsh_B9	0.74	-5.23 (0.95)	0.24	-70.15 (0.30)	1.56	105.99 (2.06)	0.63	-43.21 (0.57)	0.11
MA821Q + pLKOsh_B9	1.05	34.23 (1.34)	0.83	3.17 (1.03)	1.29	69.85 (1.70)	0.91	-17.90 (0.82)	0.60
MA821Qsh_B1	0.71	-9.60 (0.90)	0.93	15.56 (1.16)	1.10	44.49 (1.44)	0.88	-20.63 (0.79)	6.43
MA821QmiR_B1	0.82	4.77 (1.05)	0.86	7.09 (1.07)	1.07	40.68 (1.41)	0.85	-23.01 (0.77)	2.91
pLKOsh_B1	0.91	16.23 (1.16)	0.66	-18.19 (0.82)	1.01	33.18 (1.33)	0.74	-32.72 (0.67)	0.56
MA821Q + pLKOsh_B1	0.66	-16.35 (0.84)	0.07	-91.46 (0.09)	1.56	105.65 (2.06)	0.30	-73.21 (0.27)	16.15

C

	relative gene expression								VCN
	HBG		BCL11A_(B9)		BCL11A_(B1)		total HBB		
	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	
control	0.36		0.57		0.80		1.01		0.06
MA821	0.72	101.16 (2.01)	1.25	119.98 (2.20)	0.87	7.97 (1.08)	1.94	92.56 (1.93)	1.59
MA821Q	0.70	95.79 (1.96)	1.00	75.56 (1.76)	0.89	11.26 (1.11)	1.39	37.64 (1.38)	2.02
MA821QmiR-Scr	0.38	5.74 (1.06)	0.55	-3.30 (0.97)	0.75	-6.36 (0.94)	0.89	-11.86 (0.88)	1.45
pLKOsh_Scr	0.24	-32.00 (0.68)	0.29	-48.75 (0.51)	0.76	-5.51 (0.94)	0.55	-44.96 (0.55)	9.84
MA821Qsh_B9	0.54	49.94 (1.50)	1.02	79.79 (1.80)	0.86	6.82 (1.07)	1.16	15.54 (1.16)	1.13
MA821QmiR_B9	0.93	159.15 (2.59)	1.60	181.79 (2.82)	0.79	-1.12 (0.99)	2.23	120.96 (2.21)	1.08
pLKOsh_B9	0.94	163.39 (2.63)	0.50	-11.64 (0.88)	0.92	14.30 (1.14)	0.58	-42.40 (0.58)	4.03
MA821Q + pLKOsh_B9	5.50	1438.01 (15.38)	0.30	-47.45 (0.53)	0.94	17.53 (1.18)	0.67	-33.22 (0.67)	18.62
MA821Qsh_B1	0.34	-3.90 (0.96)	0.71	24.43 (1.24)	1.33	65.69 (1.66)	0.72	-28.13 (0.72)	1.53
MA821QmiR_B1	0.32	-11.46 (0.89)	0.60	5.21 (1.05)	0.87	8.75 (1.09)	0.53	-47.45 (0.53)	0.29
pLKOsh_B1	0.41	14.84 (1.15)	0.32	-43.42 (0.57)	0.89	11.49 (1.11)	0.53	-47.17 (0.53)	3.47
MA821Q + pLKOsh_B1	0.64	80.33 (1.80)	0.16	-71.59 (0.28)	0.76	-5.41 (0.95)	0.41	-58.90 (0.41)	18.33

Figure 3.5.13. RT-qPCR analysis of mRNA expression in transduced primary erythroid cultures. $CD34^+$ cells from **(A, C)** IVSI-110/N and **(B)** IVSI-110/IVSI-110 donors were transduced with various LVs as indicated, and analysed for HBG, BCL11A (regions surrounding the B9 and B1 recognition sites), and total HBB (endogenous and vector-derived) mRNA levels by RT-qPCR after 5–10 days in Phase-II medium. Values are normalized by *GAPDH* expression. Each table shows relative gene expression, the percentage change (Δ%) and fold change in expression between the transduced and the control cultures, as well as the average cell VCN in bulk samples. Data are presented for single transductions from three independent experiments.



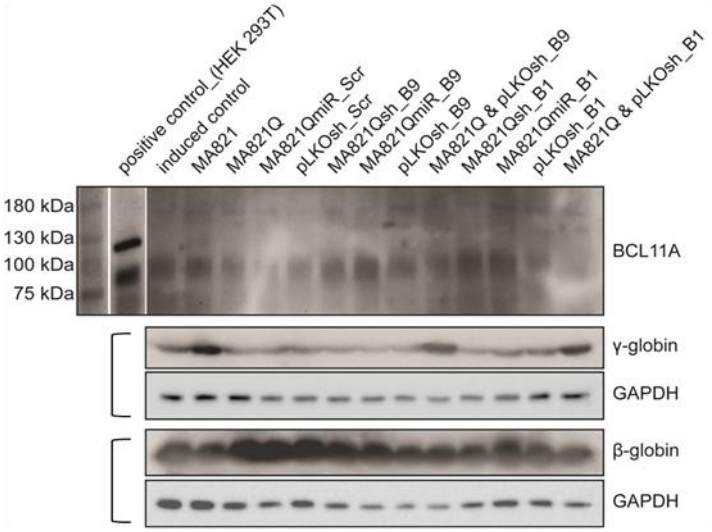
ID	sample	relative gene expression							
		total HBB		HBG		BCL11A_(B9)		BCL11A_(B1)	
		fold change (±SD)	Δ%	fold change (±SD)	Δ%	fold change (±SD)	Δ%	fold change (±SD)	Δ%
1	control	1.00±0.00		1.00±0.00		1.00±0.00		1.00±0.00	
2	MA821	1.29±0.61	28.72	1.31±0.73	30.87	1.48±0.76	47.73	1.09±0.02	9.36
3	MA821Q	0.71±0.15	-29.25	0.92±0.32	-7.69	1.65±0.16	64.58	1.06±0.08	5.96
4	MA821QmiR-Scr	0.80±0.12	-20.10	1.12±0.09	12.03	1.06±0.13	5.87	0.97±0.05	-2.52
5	pLKOsh_Scr	0.88±0.32	-12.38	0.78±0.14	-21.78	0.42±0.13	-57.88	0.93±0.02	-6.82
6	MA821Qsh_B9	1.03±0.13	2.59	1.43±0.10	42.64	0.81±0.22	-19.00	1.06±0.01	6.15
7	MA821QmiR_B9	0.39±0.53	-61.49	1.25±1.30	25.17	0.23±0.32	-76.57	0.98±0.01	-2.17
8	pLKOsh_B9	0.57±0.01	-42.81	0.47±0.67	-52.57	0.15±0.21	-84.91	1.17±0.04	16.92
9	MA821Q + pLKOsh_B9	0.74±0.11	-25.56	1.18±0.24	17.59	0.52±0.51	-47.78	1.09±0.12	9.29
10	MA821Qsh_B1	0.76±0.05	-24.38	0.94±0.03	-6.25	1.20±0.06	20.00	1.55±0.15	55.09
11	MA821QmiR_B1	0.65±0.17	-35.23	0.97±0.11	-3.35	1.06±0.01	6.15	1.25±0.23	24.72
12	pLKOsh_B1	0.60±0.10	-39.95	1.16±0.01	15.54	0.69±0.18	-30.81	1.22±0.15	22.34
13	MA821Q + pLKOsh_B1	0.34±0.10	-66.06	1.32±0.68	31.99	0.18±0.14	-81.53	1.50±0.79	50.12

Figure 3.5.14. RT-qPCR analysis of mRNA expression in transduced primary erythroid cultures. $CD34^+$ cells from IVSI-110/N and IVSI-110/IVSI-110 donors were transduced with various LVs as indicated, and analysed for **A)** HBB (endogenous and vector-derived), **B)** HBG and **C**–**D)** BCL11A (regions surrounding the B9 and B1 recognition sites) mRNA levels by RT-qPCR after 5–10 days in Phase-II medium. Values are calibrated with *GAPDH* and normalised to the control samples as 1.0, for comparability across experiments. Values are the mean ± SD relative gene expression of 2–3 cultures/LV from different experiments, following outlier detection and removal. The table shows the percentage change (Δ%) in expression between the transduced and the control cultures. Analysis was performed by one-way ANOVA and values were compared within each gene group by Tukey's post-hoc test (*, p<0.05).

Western blotting was performed to assess BCL11A knockdown, *HBG* reactivation and vector-derived *HBB* expression at the protein level. A representative experiment is shown in **Figure 3.5.15**. Analysis was performed with samples derived from transduced $CD34^+$ cells isolated from blood of a donor with thalassaemia trait (IVSI-110/N). The experimental samples showed a comparable average copy number per cell for all vectors in bulk cell populations in the range of 0.15–5.23 (mean: 1.41; median: 0.83). The BCL11A immunoblot revealed a line of bands of variable intensity at a position around 100 kDa, which could correspond to the large BCL11A isoforms. The band signal was intensified using *Photoshop CS6* software (applying contrast changes equally to the entire image) as the immunoblot exhibited a high level of background and a low signal-to-noise ratio, which in turn also impeded effective background correction of measurements. The amount of protein loaded on the gel was equivalent to whole cell extract from $0.8x10^6$ cells. Quantification of banding intensity was performed by *Fiji* software and the percentage change was determined relative to the protein band intensity of the non-transduced control cells (expression ratio set to a default value of 1). While expression analysis of BCL11A at the protein level did not corroborate mRNA data, Western data for BCL11A received little analytical credibility owing to signal obscurity on the immunoblot. The efficiency of U6-driven and intron-derived shRNA production was thus assessed via re-activation of *HBG* expression. We detected sharp HBG signals on the immunoblot which, in same-sample comparison of protein and mRNA levels after normalization to the GAPDH signal, produced varying protein-to-mRNA ratios. In particular, vectors encoding the B9 hairpin produced heightened HBG expression at the mRNA level (>1) which was dramatically reduced at the protein level (<1). Lastly, quantification of total HBB protein expression was straightforward, owing to strong band intensities even at low exposure times of 2 seconds. For HBG and HBB immunoblots we loaded protein amounts equivalent to whole cell extract from $0.4x10^6$ cells. In response to technical problems (e.g., uneven/speckled background and signal saturation), we would have had to proceed with several trials to determine the optimum conditions for these samples (of differentiated $CD34^+$ cells). As the Western blot is a time consuming procedure, we instead employed an optimized protocol based on RP-HPLC with a rapid 14 minute gradient elution to determine levels of different globin chains.

The RP-HPLC protocol was tested using samples of differentiation-phase $CD34^+$ cell extracts from normal (N/N), IVSI-110/N, and IVSI-110/IVSI-110 donors. The choice of samples was in line with our interest to investigate the basal levels of HBB expression in different β-thalassaemia genotypes and assess the feasibility/sensitivity of the technique to detect variation in HBB output with the severity of the β-thalassaemia genotype. Data are presented in **Figure 3.5.16**. The chromatographic peaks were sharp and symmetrical with low background/baseline level. In order to assign peaks by elution time, we conducted calibration assays with samples of known composition under identical analytical conditions. This included blood from a normal adult donor (for HBB and HBA chains) and cord blood (for HBG and HBA chains). Peaks were assigned by comparing the retention factor of the peak in the chromatograms and shown to elute in the order of haeme, pre-HBB, HBB, HBD, HBA, HBG2 and HBG1 chains (**Figure 3.5.16A**). This study did not include known standard solutions for determining the amount of globin chains in the samples. Instead, the 'Area%' function was used to calculate the relative amounts of proteins. We detected comparable Area% values for both haeme and HBA chains across all sample solutions and used these as internal standards to produce an expression ratio for the amount of each protein in the samples (**Figure 3.5.16B**). We observed a decreasing HBB expression ratio with increasing severity of the β-thalassaemia genotype. Compared to the normal sample (N/N), we found an 11% and 64% reduction in expression ratio (normalized to haeme) for IVSI-110/N and IVSI-110/IVSI-110 samples, respectively. Similarly, we detected a 42% and 72% reduction in expression ratio (normalized to HBA) for IVSI-110/N and IVSI-110/IVSI-110 samples, respectively (**Figure 3.5.16C**). Moreover, following on from reports that culture conditions induce expression of *HBG* in adult erythroid progenitor cells, we also measured the expression ratio of HBG1 and HBG2 chains. Compared to the adult normal blood sample, our results showed an overall increase in the HBG output and reported a 934%, 300%, and 1036% higher expression ratio (normalized to HBA) for normal N/N, IVSI-110/N and IVSI-110/IVSI-110 samples, respectively (**Figure 3.5.16C**). We concluded that the RP-HPLC was highly efficient at detecting changes in globin chain expression and decided to use RP-HPLC in place of Western blot protocols for protein quantification analyses.

A



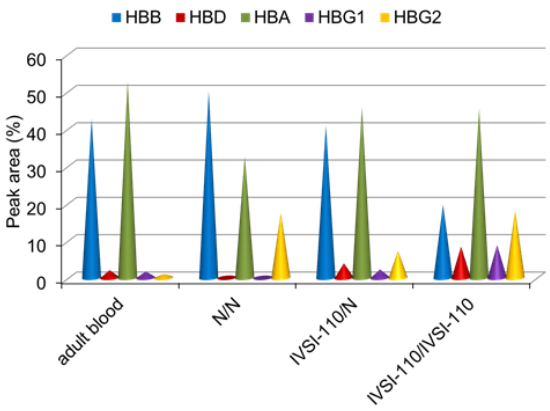
	HBB protein expression *Intensity (normalized to GAPDH)*			HBG protein expression *Intensity (normalized to GAPDH)*			BCL11A protein expression *Intensity (normalized to GAPDH)*		
	normalized ratio	expression ratio (*n*-fold)	Δ%	normalized ratio	expression ratio (*n*-fold)	Δ%	normalized ratio	expression ratio (*n*-fold)	Δ%
control	0.656	1.00		1.758	1.00		0.662	1.00	
MA821	1.058	1.61	61.25	4.010	2.28	128.06	0.742	1.12	12.04
MA821Q	8.710	13.27	1227.14	1.139	0.65	-35.21	0.244	0.37	-63.16
MA821QmiR-Scr	14.725	22.44	2143.55	1.531	0.87	-12.91	0.072	0.11	-89.05
pLKOsh_Scr	7.317	11.15	1014.87	1.370	0.78	-22.08	0.132	0.20	-80.13
MA821Qsh_B9	5.824	8.87	787.42	1.038	0.59	-40.96	0.206	0.31	-68.91
MA821QmiR_B9	8.747	13.33	1232.70	0.902	0.51	-48.70	0.295	0.45	-55.48
pLKOsh_B9	7.685	11.71	1070.96	1.606	0.91	-8.63	0.966	1.46	45.97
MA821Q + pLKOsh_B9	6.080	9.26	826.38	6.069	3.45	245.17	0.940	1.42	42.09
MA821Qsh_B1	2.362	3.60	259.89	1.306	0.74	-25.74	1.431	2.16	116.17
MA821QmiR_B1	1.710	2.61	160.54	1.275	0.73	-27.46	0.684	1.03	3.32
pLKOsh_B1	1.443	2.20	119.87	1.434	0.82	-18.47	-0.001	0.00	
MA821Q + pLKOsh_B1	1.342	2.04	104.48	2.413	1.37	37.25	0.013	0.02	-98.00

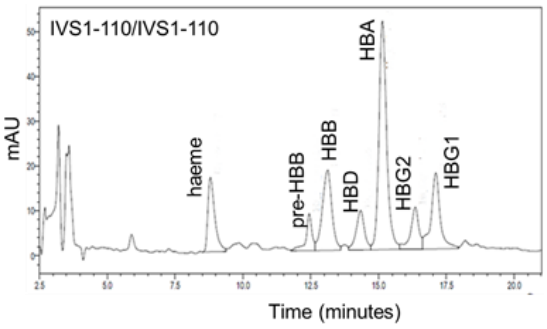
B

	relative gene expression								VCN
	HBG		BCL11A_(B9)		BCL11A_(B1)		total HBB		
	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	ratio	Δ% (fold change)	
control	0.36		0.57		0.80		1.01		0.06
MA821	0.72	101.16 (2.01)	1.25	119.98 (2.20)	0.87	7.97 (1.08)	1.94	92.56 (1.93)	1.59
MA821Q	0.70	95.79 (1.96)	1.00	75.56 (1.76)	0.89	11.26 (1.11)	1.39	37.64 (1.38)	2.02
MA821QmiR-Scr	0.38	5.74 (1.06)	0.55	-3.30 (0.97)	0.75	-6.36 (0.94)	0.89	-11.86 (0.88)	1.45
pLKOsh_Scr	0.24	-32.00 (0.68)	0.29	-48.75 (0.51)	0.76	-5.51 (0.94)	0.55	-44.96 (0.55)	9.84
MA821Qsh_B9	0.54	49.94 (1.50)	1.02	79.79 (1.80)	0.86	6.82 (1.07)	1.16	15.54 (1.16)	1.13
MA821QmiR_B9	0.93	159.15 (2.59)	1.60	181.79 (2.82)	0.79	-1.12 (0.99)	2.23	120.96 (2.21)	1.08
pLKOsh_B9	0.94	163.39 (2.63)	0.50	-11.64 (0.88)	0.92	14.30 (1.14)	0.58	-42.40 (0.58)	4.03
MA821Q + pLKOsh_B9	5.50	1438.01 (15.38)	0.30	-47.45 (0.53)	0.94	17.53 (1.18)	0.67	-33.22 (0.67)	18.62
MA821Qsh_B1	0.34	-3.90 (0.96)	0.71	24.43 (1.24)	1.33	65.69 (1.66)	0.72	-28.13 (0.72)	1.53
MA821QmiR_B1	0.32	-11.46 (0.89)	0.60	5.21 (1.05)	0.87	8.75 (1.09)	0.53	-47.45 (0.53)	0.29
pLKOsh_B1	0.41	14.84 (1.15)	0.32	-43.42 (0.57)	0.89	11.49 (1.11)	0.53	-47.17 (0.53)	3.47
MA821Q + pLKOsh_B1	0.64	80.33 (1.80)	0.16	-71.59 (0.28)	0.76	-5.41 (0.95)	0.41	-58.90 (0.41)	18.33

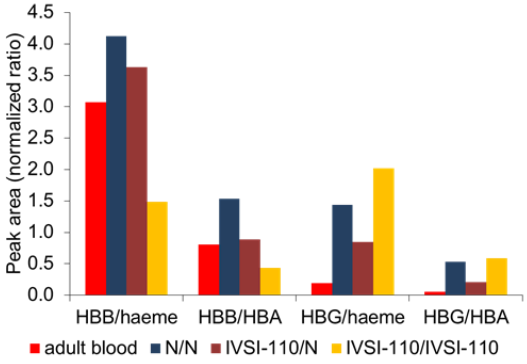
Figure 3.5.15. Western blot analysis of globin chain expression in primary erythroid cultures. $CD34^+$ cells (IVSI-110/N) were transduced with different LVs, as indicated, and analysed for globin chain expression after 7 days in Phase-II medium. **A)** Western blots from whole cell extract were analysed with antibodies against BCL11A (*Santa Cruz*), HBG and HBB. GAPDH was used as internal control. Molecular weight of BCL11A isoforms XL/L/S/XS: 125/100/35/25 kDa. **B)** RT-qPCR assays detected mRNAs for HBG, BCL11A (regions surrounding the B9 and B1 recognition sites) and total HBB (endogenous and vector-derives), as indicated. Shown are relative quantities normalised to *GAPDH* expression.

A



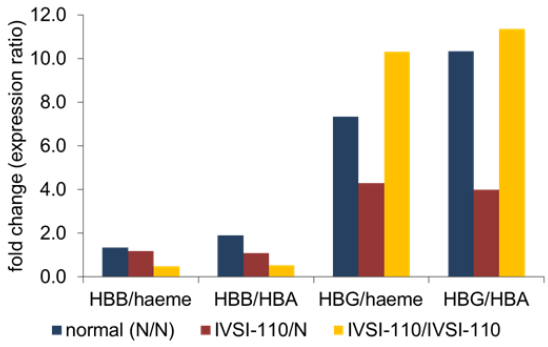


B



	normalized peak area (ratio)			
	HBB/haeme	**HBB/HBA**	**HBG/haeme**	**HBG/HBA**
adult blood	3.07	0.81	0.20	0.05
(+/+)	4.13	1.53	1.44	0.54
IVS1-110/+	3.63	0.89	0.84	0.21
IVS1-110/IVS1-110	1.49	0.43	2.02	0.59

C



	HBB/haeme		**HBB/HBA**		**HBG/haeme**		**HBG/HBA**	
	fold change	Δ%	fold change	Δ%	fold change	Δ%	fold change	Δ%
adult blood	1.00		1.00		1.00		1.00	
(+/+)	1.34		1.89		7.35	634.93	10.34	934.21
IVS1-110/+	1.18	-11.95	1.10	-41.90	4.30	330.48	4.00	299.71
IVS1-110/IVS1-110	0.48	-63.96	0.53	-71.79	10.31	931.48	11.36	1036.40

Figure 3.5.16. RP-HPLC analysis of globin chain expression in primary erythroid cultures. $CD34^+$ cells from normal (N/N), IVSI-110/N, and IVSI-110/IVSI-110 donors were analysed for HBA, HBB and HBG1/HBG2 chains by RP-HPLC after 9 days in Phase-II medium. **A)** The histogram reports the area of each peak in the chromatogram as a percentage (%) of the total area of all peaks. The plot of chromatographic peaks on the right corresponds to the IVSI-110/IVSI-110 sample and is a representative chromatogram for all other samples. **B)** The histogram reports a ratio value of the peak area of the HBB and HBG chains normalized to the peak area of the haeme and HBA. **C)** The histogram reports the expression ratio of HBB and HBG normalized protein amounts in each experimental condition (genotypes) relative to the control sample (adult blood), which is arbitrarily set at a value of 1. The table shows the percentage change (Δ%) in HBB expression between the IVSI-110 genotypes and the normal sample (N/N), as well as the Δ% in HBG expression between all $CD34^+$ genotypes and the control sample (adult blood).

3.5.3 Discussion

In many aspects, β-thalassaemia is a suitable candidate disease for *ex vivo* gene therapy. In order for vectors to enter a clinical trial, first, pilot runs for pre-clinical evaluations, *in vitro* and *in vivo*, are required. In this study, we sought to evaluate the pre-clinical efficacy of our LVs in human $CD34^+$ cells purified from PB, which required the establishment of a clinically relevant primary erythroid cell culture model system with ability to predict the correction of the β-thalassaemia phenotype in a large pool of erythroid cells.

In the present investigation, major technical difficulties prompted us to consider the use of three different *in vitro* liquid culture protocols. The culture system by Fibach (Protocol B) showed cell proliferation and differentiation with a starting population of light-density cells but failed to support viability with a starting sample of transduced $CD34^+$-enriched cells, while the culture system by Migliaccio (Protocol A) worked efficiently with transduced $CD34^+$ cells and was thus used for subsequent experiments. We showed that the proliferative assay efficiently supported amplification of PB-derived $CD34^+$ cells even though expansion rates would vary, depending on the individual blood donor from whom the cells had been derived. The success of the protocol was mainly limited by the small number of $CD34^+$ cells obtained from PB for large-scale experiments on all vectors. In spite of an efficient expansion, the proliferation rate was consistently low and viability was occasionally compromised especially with a starting sample of transduced $CD34^+$-enriched cells. Those cultures that survived to the end of the experimental period of culture had efficiently differentiated to erythroid cells, but the resulting cell material was insufficient to perform all necessary analyses. We speculated that the limiting cell material was the source of apparently unreliable assay results. The expansion of $CD34^+$ cells prior to transduction was not feasible since the proliferation-phase conditions also induced erythroid differentiation. To produce clinically relevant data, it is imperative that vector integration precedes the activation of differentiating factors as would be the case with any erythroid gene expressed from its natural context.

As an alternative we used the buffy coat layer from blood of healthy donors as a richer source of $CD34^+$ cells. It has been reported that the light-density cell fraction of the buffy coat layer is

richer in $CD34^+$ cells than in whole blood [462], with an average 4.0±2.3 $CD34^+$/µl (range 0.8–10.2; n=43) in buffy coat vs. 2.4±1.3 $CD34^+$/µl (range 0.8–5.9; n=28) in normal whole blood [463]. However, the buffy coat layer in our experiments also showed variable performance and will, moreover, not be available for thalasssaemic samples. Therefore and in order to set-up experimental conditions for volumes of 20 mL peripheral thalassaemic blood, a new two-phase liquid culture system was considered [180] (Protocol C). A key feature of this protocol was an initial expansion step, which would relieve the bottleneck of having too few cells for follow-up analyses. After initial success, this protocol gave persistent problems of limited expansion, spontaneous differentiation, and sudden cell death after initial experiments, despite numerous efforts of optimization and re-ordering of reagents. It was only late in the course of the PhD project that such a system was successfully established and tested to function efficiently and reproducibly in our laboratory. As experiments on their own take time with an average culture lasting 2–3 weeks from seeding to harvesting cells for analyses, only a limited amount of data was generated for each vector.

In order to achieve knockdown of the BCL11A in $CD34^+$ cells, we compared the efficiency of mRNA modulation using RNA-pol-III and RNA-pol-II promoter-driven shRNA vector backbones. Our results conclusively showed that the shRNA structure was non-functional. In support to this result, we demonstrated an efficient BCL11A knockdown by ~85% (p<0.05) using RNA-pol-III-directed BCL11A_449-shRNA and only a minor reduction of *BCL11A* expression (~19%, p>0.05) using intron-derived shRNA production. The data showed that the siRNA_(B9) duplex sequence efficiently induced potent mRNA knockdown and indicated a sub-optimal shRNA design for RNA–pol-II-mediated expression in the modified GLOBE-based vector. The BCL11A_451-shRNA hairpin structure exhibited poor performance via the use of the RNA-pol-III-based transcription system, indicating that the siRNA_(B1) duplex sequence had reduced effectiveness for gene silencing and was thus discarded for use in subsequent experiments. Moreover, we demonstrated an efficient BCL11A knockdown by ~80% (p<0.05) using the BCL11A_449-shRNA expressed within the context of the endogenous microRNA miR-30 backbone via the RNA-pol-II promoter. We have noted that at an average VCN of 1.1 per cell, the shRNA-miR construct failed to trigger potent BCL11A knockdown, an observation which was

in line with published reports of an inefficient target knockdown (>80% reduction in protein expression) with single-copy expression [228, 464]. Therefore assuming low changes in expression, especially with gene regulation by shRNAmiR constructs, more biological samples are needed in order to reduce variability of expression values between measurements.

The repression of *BCL11A* has been shown to induce the expression of *HBG*. On the contrary, we reported an unchanged or lower *HBG* expression (p>0.05) in samples transduced with vectors which triggered potent BCL11A knockdown. In addition, these samples exhibited a lower total HBB output compared to non-transduced control cells, which was nonetheless relatively lower than the HBG output. Given the position of *HBG1/HBG2* in the *HBB* locus relative to the LCR, the looping model would favour interaction of this master regulatory element with the *HBG* promoter to form an active chromatin hub, thus enhancing *HBG* expression [95, 99]. Of note, the expression level of genes introduced by transduction with high VCN (e.g.: experiment in **Figure 3.5.13)** was possibly deregulated with respect to the physiological level of expression. This reasoning could possibly explain the differential gene expression detected using the pLKOsh_Scr vector, which exhibited an increasing drop in expression with increasing VCN. In support of this, when the same shRNA scramble sequence was expressed from within the miR30 scaffold at low VCN, the off-target phenotype was not observed.

In addition, we did not detect a significant difference in the total HBB output of LV-treated cultures suggesting that differential expression of *HBB* was mainly due to chance variation. In this study, individual subjects were carriers of the IVSI-110 (G→A) splicing mutation, which creates an alternative acceptor AG site 19 bp 5' to the normal acceptor. Since the alternative splice site is preferentially used in 80%–90% of the transcripts, only 10%–20% of the transcripts are normally spliced [57] and detected (to saturation nevertheless) on immunoblots. The HBB mRNA is unusually stable (half-life varies from 10–24 h) since translation is required for up to 3 days after transcriptional arrest and subsequent enucleation in order to sustain synthesis of globin proteins [465]. Therefore it could be possible that stable mRNAs translated over a long time would increase the level of full-length protein accumulating in the cells. Nevertheless, using an optimized protocol based on RP-HPLC, we successfully detected variation in HBB protein expression between heterozygote and homozygote subjects for the IVSI-110 mutation. In

addition, the RP-HPLC has an overall smaller requirement for primary erythroid cells ($0.5x10^6$ cells per reaction) than the Western blot ($0.3–1x10^6$ cells per reaction) since it facilitates the analysis of HBA, HBB, and HBG globins in a single reaction. We therefore propose to use this RP-HPLC protocol in place of Western blot assays for quantification of protein expression in future experiments.

In conclusion, we successfully established a reproducible and predictive *in vitro* erythroid model that allows evaluation of β-thalassaemia erythroid precursor cells transduced with our GT LVs. Initial observations were promising but not reproduced in two subsequent experiments. Thus more experiments are needed to infer a clear trend for application of our GT vectors and to draw conclusions as to their functionality with (statistical) certainty. Overall, this study provided strong evidence to refute the functionality of the intron-derived shRNA hairpin, whilst it provided encouraging data to assume potent target knockdown using the shRNA-miR hairpin design.

4.0 Final Discussion

4.1 Rationale and vector design

The key aim of this project was the development of improved LVs for the GT treatment of β-thalassaemia (and sickle cell disease). The strategy adopted aimed to supplement vector-mediated *HBB* expression with targeted induction of endogenous *HBG*. Based on the GLOBE vector platform, with its reduced LCR and high vector titres at high expression levels [156, 188], LVs were engineered to co-express shRNAs against inhibitors of *HBG* expression, namely, BCL11A, SOX6 and KLF1. Synthetic shRNA-encoding sequences were inserted at the junction point of the internal IVS-II deletion in the GLOBE vector. This site of insertion is the logical position for the re-introduction of additional sequences and is close to the position successfully used by others to express an anti-sickling-β-globin-specific shRNA from a recombinant *HBG* cassette [224]. LVs expressing GFP-specific shRNAs were produced to quickly assess the functionality of the shRNA design and cloning site as a means to verify the functionality of the vector concept. Analyses in GFP-transgenic cell lines suggested that the initial shRNA design was problematic, and a search of the literature suggested that the design of RNA-pol-II-mediated expression of the shRNA in the modified GLOBE-derived LVs was sub-optimal. Through attempts to recapitulate *in silico* the cloning steps from the original study on which our own design was based, it appeared that the description of sequences and components in the latter was incomplete (**Figure 1.14**) [224]. In the light of a substantial body of independent work on the expression of shRNAs from RNA-pol-II promoters it became apparent, moreover, that flanking sequences would be required to allow efficient processing of the intron-encoded shRNA in GLOBE-derived shRNA LVs [304, 305, 308]. Therefore a new generation of shRNA-based vectors was produced, in which the shRNA was flanked by miRNA miR30 target sequences that should result in excision and improved efficiency of shRNA processing. These new vectors, called shRNA-miRs, were introduced late in the project and therefore produced only for the BCL11A-specific shRNA and control scramble-shRNA sequences used within simple hairpin structures. BCL11A has been recognized to be one of the most important regulators of the fetal-to-adult Hb switch and a core transcription factor within erythroid multiprotein complexes guiding HbF silencing in mature erythroblasts. BCL11A therefore represented the most promising endogenous target for an initial assessment of vector-mediated HBG reactivation and HBB

expression using the shRNA-miR design. Additionally, independent efforts by the Williams group [466] indicate that HBG induction through BCL11A knockdown driven by the *HBB* promoter is a working concept but that sequence modifications may be required to adapt shRNAs from RNA-pol-III-based expression for use in RNA-pol-II-based expression vectors.

4.2 Erythroid models

A prerequisite for confirming the achievement of the project aim, i.e. of an improvement of the extant GLOBE vector, was the functional characterisation of the novel LVs in primary human erythroid progenitors. Difficulties encountered in establishing an informative primary cell culture model system led to the investigation of alternative model systems early on in the project, in parallel to on-going efforts to establish an *in vitro* primary erythroid liquid culture system (**section 4.3**).

Early experiments were conducted using shRNA that would target *GFP* reporter expression to quickly establish proof of principle for vector design. Initially, BM Lin^- stem cells from an inbred GFP-transgenic mouse were harvested, transduced and cultured in methylcellulose-based media for two weeks to allow their development into discrete colonies. Analysis for shRNA-mediated GFP knockdown in individual colonies by flow cytometry was uninformative, since the number of cells per colony was too small for signal detection above flow cytometer background noise and for additional assessment of VCN to correlate transduction of clones with GFP expression. The re-infusion of transduced cells into either GFP-transgenic or wild-type myeloablated mice was considered, but would have been flawed by the only available flow-cytometry marker of chimerism (GFP) at the same time also being the target of RNAi, thus impeding the discrimination of donor and recipient cells in injected mice.

As a consequence and to ensure an abundance of material for analysis, GFP-transgenic HEL human erythroleukemia cell line clones were produced by transduction with the PGK-GFP LV and by two rounds of isolation of single clones. After establishment of optimal conditions for erythroid differentiation and transduction of these cells, it became apparent that differentiation alone profoundly reduced *GFP* expression (>90%) and would therefore mask any silencing of GFP mediated by GLOBE-derivatives, expression of which depends on erythroid differentiation.

We speculated that treatment with the erythroid inducer haemin could have altered the chromatin domain of the integration site and/or deregulated the activity of the PGK promoter, thus accounting for variegating GFP silencing. Assessment of vectors in undifferentiated HEL cells instead did not provide clear information on shRNA functionality, with low levels of vector-derived expression as a feasible explanation for non-detection of GFP silencing. The particular batch of HEL cells in question, however, expressed vector-derived *HBB* mRNA even without erythroid induction and thus provided the first proof that the LV-encoded *HBB* transgene in GLOBE-derived vectors was functional.

In order to assess shRNA processing nevertheless, the murine erythroleukaemia (MEL) cell line was next employed as a second alternative to primary human cells, once more based on stable transduction with the PGK-GFP LV to allow assessment of GFP knockdown. Seven-day induction to terminal erythroid differentiation and subsequent analysis for GFP knockdown revealed elevated expression of vector-derived *HBB* and absence of vector-independent PGK-GFP down-regulation in control cells. However, no RNAi activity was detectable by RT-qPCR for RNAi target sequences, casting doubt over the chosen LV-derived shRNA expression strategy. A simple hairpin design driven from the RNA-pol-III U6 promoter mediated potent GFP knockdown (>70%) and thus demonstrated the functionality of the shRNA stem and loop sequences employed in GLOBE-derived vectors. This narrowed failure to detect RNAi activity from the MA821Q-derived vector down to absent or reduced processing of functional shRNA sequences from the vector-encoded LV HBB intron 2.

Turning away from reporter-gene detection and towards the detetion of disease-relevant endogenes, we established the human myelogenous leukaemia cell line K562, with embryonic/fetal globin gene expression, as a third alternative to primary human cells for vector assessment. In K562 cells, the *BCL11A* endogene was targeted for knockdown, which, however, gave persistent background bands on immunoblots, and which by RT-qPCR gave contradictory results for primer-probe combinations recognising different parts of the *BCL11A* mRNA. Abandoning K562 cells as a model for BCL11A knockdown, we applied the same BCL11A-specific vectors and corresponding detection by immunoblots and RT-qPCR in HEL cells. This revealed the latter, which express the large and HBG-repressing BCL11A isoforms,

as a more suitable platform for the analysis of vector-mediated BCL11A knockdown. However, undetectable shRNA activity from the U6 promoter-based control vectors, which expressed publically available and validated shRNA, also prompted the dismissal in HEL cells of BCL11A-targetting RNAi. Subsequent quantitative studies for *BCL11A* expression across different cell lines confirmed the presence of the large isoforms of BCL11A in non-erythroid HEK 293T cells, which showed a profound reduction in *BCL11A* expression (>50%) following treatment with U6-driven shRNAs. Overall, these analyses in cell lines have indicated that shRNA expression was absent using the original globin gene-based LV design, where stem-loop hairpin sequences are inserted within the second intron of *HBB*.

Differential effects on BCL11A and HBG expression by BCL11A shRNAs expressed under the constitutive U6 promoter in different cell lines and using identical methods for RNAi delivery and detection were unexpected and, as yet, have no clear explanation. Differences might be rooted in different genome architectures of the three immortal cell lines investigated. The embryonic-fetal globin expressing K562 and HEL cell lines have a lymphoblast morphology characterized by several chromosomal abnormalities [250, 251], while the embryonic kidney HEK 293T cell line with fibroblast morphology, albeit of normal origin, has acquired a highly aberrant karyotype following cellular immortalization [253]. Moreover, BCL11A has been linked to cellular proliferation and cancers [467], and HEL and MEL originating from leukaemic and HEK293 from normal cells might modulate BCL11A-related signalling pathways, also it is unclear how this could affect RNAi and mRNA detection for BCL11A differentially.

4.3 Primary erythroid cultures

Functional characterization of LVs *in vitro* with erythroid liquid cultures required more sample material than was originally available through thalassaemic PB samples that are received thrice weekly by our laboratory and represented a readily available source of an albeit limited number of human $CD34^+$ cells. These samples also represent a valuable and diverse resource for the establishment of a biobank of thalassaemic HSPCs. Establishment of a biobank as well as pooling of same-patient samples over time for larger-scale experiments required cryopreservation, which prompted us to evaluate different PBMC isolation protocols (**sections**

2.9.2.1–2.9.2.3) and subsequent cell viability and differentiation potential in CFC assays. Direct controlled-rate freezing of PB in 10% DMSO gave poor cell viability upon thawing and was thus dismissed as an option for cryopreservation. Of a range of methods tested subsequently, simple ACK lysis of RBCs or density-gradient centrifugation, combined with cryopreservation of the resulting leukocytes in 90% and 50% FBS-supplemented freezing media, respectively, gave satisfactory cell recovery, viability and performance in scaled-down clonogenic assays. This study is novel in terms of comparing extraction protocols and cryostorage conditions for optimal recovery, viability and lineage-potential of PB-isolated HSPCs, of relevance not just to our own group, but to any laboratory with access to small-volume diagnostic samples and the corresponding informed consent for cryostorage and experimental use of samples. Pending further tests and comparison, Lymphoprep+(50%FBS) has been adopted by our laboratory for the long-term storage of incoming small-volume diagnostic samples.

$CD34^+$ cells, the target for clinical and most preclinical GT approaches targeting the haematopoietic system, were isolated from the PBMC fraction by MACS selection. A first culture system used (Protocol A), which supports growth of peripheral $CD34^+$ cells and commitment to erythropoiesis in the presence of stimulators of the glucocorticoid receptor and a variety of growth factors (SCF, IL3, and EPO) [188, 335], was limited by the small number of $CD34^+$ cells obtained from PB. The system showed limited expansion and reduced viability of cells especially with a starting sample of LV transduced $CD34^+$-enriched cells. Persistent low performance of this and of a second culture system (Protocol B) [454] that amplified erythroid-committed progenitors in an EPO-independent phase and then induced differentiation into mature erythroblasts by exposure to EPO, prompted adoption of a third, two-phase liquid culture system [180]. This system (Protocol C) was already used at the time by collaborating institutions, was to be adopted by our group for a collaborative project and was known to work efficiently with similar experimental parameters, such as lentiviral transduction with HBB-expressing vectors and use of naïve thalassaemic PB samples as source material. Initial variation in cell yields, also during the characterisation of my vectors as reported here, have been overcome by standardisation of locally available reagents and methods, and the protocol

has now been established in our group as a standard method for high-yield HSPC biobanking for starting sample sizes of 25 mL or more.

Initial experiments using Protocol C aimed to measure MA821Q-derived *HBB* expression compared to the base GLOBE LV. Evaluation in IVSI-110/N $CD34^+$ cells from a single experiment, which achieved comparable VCN of 1–2, suggested more efficient gene delivery with the original GLOBE vector. The modified GLOBE-T87Q vector contains additional IVS-II sequences, which constitute the BsrG*I* and BsiW*I* restriction sites. As already described in **section 3.1.2**, the artificial insertion of the BsiW*I* site led to the unintentional modification of the last polypyrimidine tract base from a pyrimidine (cytosine; C) to a purine (guanine; G). In the absence of direct evidence from site-directed mutagenesis studies to suggest otherwise, we speculated that the nucleotide change would not impinge on lariat formation and/or splicing efficiency. However, in the light of our own data and with data on the mutation of adjacent sites leading to lowered HBB expression after all, a direct comparison of a HBB^{T87Q}-encoding vector with and without the BsiW*I*-related nucleotide changes is indicated to confirm or refute our intial assumption. Thus at present we do not know the cause of the lower levels of expression from GLOBE-T87Q compared to GLOBE.

Preliminary analyses in primary cells demonstrated that intron-encoded shRNAs showed no RNAi activity in accordance with findings in the cell line systems. Testing RNAi efficiency with U6 promoter-based vectors revealed hairpin potency only for the BCL11A-specific siRNA_(B9), the sequence of which in the context of the miR30 scaffold and driven by the *HBB* promoter produced effective, albeit inconsistent, BCL11A protein knockdown. This raised the possibility that the miR scaffold design is sub-optimal for small RNA processing and consistent knockdown. Moreover, where increased HBG expression was achieved, this did not correlate with BCL11A knockdown, in contrast to published reports [63, 119, 233, 466]. Reports of a decline of the expression of the large BCL11A isoforms in mature erythroblasts [63, 369] raise concern over a possible masking of BCL11A RNAi by developmental down-regulation and over the relevance of our end-point analyses of BCL11A protein levels for HBG induction at earlier timepoints. The reported peak expression of large BCL11A isoforms in proerythroblasts and basophilic normoblasts followed by gradual depletion with ongoing maturation [63, 369] equally

raised concern over the temporal regulation of BCL11A-shRNA expression from the *HBB*-promoter and over the timeliness of BCL11A knockdown for HBG induction. However, a similar approach has lately also been pursued by the Williams group with significant efficiency in the induction of HBG in preliminary reports [233, 466], so that these latter concerns appear to be unfounded. The latter work, however, also and unexpectedly indicated that expression from miRNA scaffolds leads to a shift in the processing of shRNA sequences compared to expression from the U6 promoter and might thus require a modification of our shRNA-encoding sequences to produce intact siRNAs and effective RNAi.

In conclusion, in the work described in this thesis we have established numerous methods inherent to *ex vivo* GT applications in a novel research environment and have achieved, amongst other things, high-efficiency HSPC expansion from PB, biobanking of low-volume PB samples, high-titre LV production, qPCR, RT-qPCR, RP-HPLC and immunoblot detection as key methods for GT of β-thalassaemia. With the data produced using these methods, we demonstrated that expression of shRNAs from simple stem-loop structures inserted within an intronic sequence under control of the RNA-pol-II β-globin promoter does not result in effective shRNA production and target sequence knockdown. The development of a novel strategy using miR30 target site-adapted shRNAs, although promising, has not, as yet, been functionally confirmed. The present study gives a basis for future work on improved LV GT vectors for β-thalassaemia that combine vector-derived HBB expression and induction of endogenous HBG for curative Hb levels at low VCN.

5.0 Future Work

5.1 Functional characterization of LVs in CD34$^+$ cells

A key aim of this study was to assess the therapeutic potential of the modified GLOBE-T87Q vector and its RNAi-based permutations relative to the original GLOBE vector [188]. The latter has been functionally characterized in β-thalassaemia erythroid cells to correct hallmark features of the β-thalassaemia phenotype and provide an efficient therapy with relatively low VCN [188]. In this study, in spite of establishing basic vector-derived *HBB* expression in cultures of CD34$^+$ cells, as yet, comparable analyses have been unfruitful owing to large differences in VCN between vectors across different experiments. Further work would entail running replicate transductions at an estimated VCN of 1, followed by quantification of HBB and HBG protein levels by RP-HPLC in order to measure i) vector-derived *HBB* expression compared to base vector, ii) impact of concomitant shRNAmiR expression on vector-derived HBB levels, and iii) competition between vector-derived *HBB* and endogenous *HBG* genes.

Furthermore, in order to assess the performance of intron-derived shRNA production, erythroid liquid cultures should be set-up and analysed at different stages of erythropoiesis to determine the expression profile of the shorter and larger isoforms of BCL11A in adult erythroid cells. Consequently, RNAi-mediated silencing of BCL11A will be measured at the erythroblast stage where BCL11A-XL is abundantly expressed.

5.2 Optimization of RNAi expression strategy

On the grounds of difficulties to detect BCL11A knockdown in adult erythroid cells even for the U6-driven shRNAs, we recommend development of vectors carrying shRNA-miR cassettes encoding the sequence targeting GFP in order to provide proof-of-principle for vector design. The performance of the shRNA-miR vector driven by RNA-pol-III U6 promoter relative to RNA-pol-II *HBB* promoter would be tested in GFP-expressing MEL cells by flow cytometry using established protocols. In case GFP silencing is not detected, instead, the following will be applied:

5.2.1 Changing the cloning site of shRNA-miRs

The shRNA-miR cassette is currently inserted in the junction point of internal IVS-II deletion in the MA821Q vector (**section 3.1.4)**. With the aim of enhancing knockdown efficiency, the cloning position of the shRNA-miR within IVS-II of *HBB* could be changed to a similar location to what has been previously described [224]. It has been shown that efficacious shRNA-miR activity can be achieved by insertion within mRNA untranslated regions (UTRs). However, published work [224, 319] where shRNA-miR cassettes were inserted within the 5'UTR or 3'UTR of the gene backbone demonstrated moderate RNAi activity with concomitant reduction in transcript transgene levels, which would be an undesirable outcome from our vectors.

5.2.2 Modification of the shRNA-miR structure

Currently, understanding of the structural requirements for effective single-copy RNAi activity is limited, restricting rational optimization of existing shRNA-miR systems [228]. Studies aiming to explore features associated with effective processing of endogenous pri-miRNAs demonstrated the need for a UG di-nucleotide sequence at the 5' basal stem and a UGUG at the 5' loop end positions for optimal performance, while the presence of an ACNNC motif 3' of the basal stem was required for pri-miRNA recognition [228, 468]. All three features were included in the shRNA-miR construct used in our study. Notably, it has recently been shown that replacing UGUG with AGUG led to improved knockdown for miR30-embedded shRNAs, posing a promising modification for our shRNA-miR design [228]. More drastic changes to the miR30 scaffold were reported in a study targeting CCR5 expression by adjusting complementary sequences and stem lengths. The miR30 loop was retained while the backbone was replaced with miR16 sequences to produce a variant with greater knockdown potency than the original miR30 [469]. During preparation of this thesis a report describing the development of improved shRNA-miR scaffolds by incorporating a 3-4bp shift in the guide strand sequence to produce mature miRNAs identical to U6-driven shRNAs was published [233, 466]. Without providing specific details of this modification, the authors reported expression of BCL11A-specific shRNA-miRs in erythroid cells under a *HBB* promoter/LCR element, resulting in increased *HBG* expression and substantial HbF production [233, 466].

Furthermore, the miR30 scaffold could be modified to incorporate the stem sequences of endogenous miRNAs, which target the transcript of choice in the erythroid compartment. Of note, the majority of endogenous miRNAs have been shown to contain putative binding sites in the 3'UTR of candidate targets for mRNA degradation and/or translational repression [470]. Particularly, miRNA-486-3p has been shown to target the BCL11A-XL 3'UTR in adult erythroid cells, contributing to elevated HbF albeit at the expense of a moderate decrease in erythroid maturation [471]. Unexpectedly, miRNA-30 was shown to target BCL11A 3'UTR in cancerous cell lines [472]. The miR30 pri-miRNA backbone is one of the most characterized and commonly used in development of therapeutic RNAi expression vectors. Replacing the miR30 backbone with that of miRNAs expressed in the erythroid compartment and/or with specificity for candidate targets will require development of multiple structural variants in order to identify the most efficient miR scaffold.

5.3 Targeted genome editing using engineered nucleases

Genome editing technology (**section 1.5.2**) can be used to facilitate permanent induction of endogenous *HBG* expression by abolishing or reducing the action of *HBG* repressors in the erythroid lineage. To this end, a recent study demonstrated NHEJ-mediated suppression of BCL11A only in erythroid cells by targeting the erythroid-restricted enhancer of *BCL11A*, although this strategy was not tested in HSPCs of β-thalassaemia patients to investigate concomitant production of HbF [123]. If successful, genetically modified HSPCs could enable permanent production of therapeutic levels of HbF, improving the β-thalassaemia phenotype. In addition to targeted genome editing [155], the CRISPR/Cas9 system was recently developed to regulate endogenous gene expression. In this case, catalytically inactive Cas9 fused to a transcriptional repression domain and targeting sites near the promoter region of targeted genes is used to silence gene expression [473].

6.0 Appendix

OPEN

Gene Therapy (2015), 1–6

www.nature.com/gt

SHORT COMMUNICATION

Measurement of lentiviral vector titre and copy number by cross-species duplex quantitative PCR

I Christodoulou[1], P Patsali[2,3], C Stephanou[2,3], M Antoniou[3], M Kleanthous[1,2,4] and CW Lederer[1,2,4]

Lentiviruses are the vectors of choice for many preclinical studies and clinical applications of gene therapy. Accurate measurement of biological vector titre before treatment is a prerequisite for vector dosing, and the calculation of vector integration sites per cell after treatment is as critical to the characterisation of modified cell products as it is to long-term follow-up and the assessment of risk and therapeutic efficiency in patients. These analyses are typically based on quantitative real-time PCR (qPCR), but as yet compromise accuracy and comparability between laboratories and experimental systems, the former by using separate simplex reactions for the detection of endogene and lentiviral sequences and the latter by designing different PCR assays for analyses in human cells and animal disease models. In this study, we validate in human and murine cells a qPCR system for the single-tube assessment of lentiviral vector copy numbers that is suitable for analyses in at least 33 different mammalian species, including human and other primates, mouse, pig, cat and domestic ruminants. The established assay combines the accuracy of single-tube quantitation by duplex qPCR with the convenience of one-off assay optimisation for cross-species analyses and with the direct comparability of lentiviral transduction efficiencies in different species.

Gene Therapy advance online publication, 23 July 2015; doi:10.1038/gt.2015.60

INTRODUCTION

Lentiviral vectors (LVs) allow curative gene therapy (GT) through genome integration, accept large payloads at high vector titre, readily transduce non-dividing cells, have a favourable safety profile compared with other integrating viral vectors and, depending on pseudotype, have wide tissue and host tropism, and thus wide-ranging applicability in the field of GT.[1] Efficacy and safety of their application are critically determined by the average vector copy number (VCN) in the corrected cell product, with more integration sites generally giving higher transgene expression levels in gene augmentation approaches, while also increasing the number of viral integration sites and thus the risk of insertional mutagenesis.[2] The desired VCN, as dictated by the required expression levels and percentage of corrected cells on the one hand and by the efficacy, long-term stability and homogeneity of transgene-derived gene expression on the other, is approximated by application of a specific number of biologically active vector particles to a known number of target cells. The actual VCN in the cell product in combination with culture and transplantation regimens, and with a possible growth advantage of corrected cells in turn determine the average VCN in the target tissues of treated patients post GT.

Initial biological vector titre, actual VCN in the cell product and post-GT VCN in the patient are usually determined by the same quantitative real-time PCR (qPCR) procedure, optimised for the respective system under study.[3] The principle of VCN determination for integrating vectors is the quantification of proviral vector genomes in extracted genomic DNA (gDNA) and their normalisation by the number of haploid host genomes against a standard curve of plasmid dilutions or of transgenic material of known copy number. The choice of LV and in particular of endogene target sequences differs considerably between laboratories, but the underlying qPCR assays almost invariably rely on (for example, TaqMan, Applied Biosystems, Waltham, MA, USA) probe technology instead of double-stranded DNA-integrating dyes, in order to minimise the detection of spurious byproducts that is common to the latter detection method. The application of labelled probes in turn opens up the possibility of duplexing LV- and endogene-specific qPCR assays labelled with different detection dyes, in order to allow same-tube detection of both sequences and with the inherent benefits of removing same-sample tube-to-tube variation and reducing reagent cost and processing time. However, to date only the Kohn group has developed a duplex protocol for LV VCN determination,[4] which has as yet not been used independently.[5–8] Strikingly and although any GT clinical trial is as a rule preceded by preclinical validation in non-human disease models, only one assay, incompatible with duplex application, has as yet been published that uses the same endogene qPCR assay for human and animal LV-GT samples.[4] Application of independent qPCR assays, however, precludes a direct comparison of VCN quantitation between species and, on the practical side, duplicates the effort and cost of assay optimisation. As an overriding dilemma, the absence of a community-wide standard quantitation assay and the habitual underdocumentation of technical details pertaining to VCN determination in outcome-focussed publications compromise the comparability of VCN and vector activity per copy between different research groups. This study therefore set out to establish

[1]Cyprus School of Molecular Medicine, Nicosia, Cyprus; [2]Department of Molecular Genetics Thalassaemia, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus and [3]King's College London, Gene Expression and Therapy Group London, UK. Correspondence: Dr CW Lederer, Department of Molecular Genetics Thalassaemia, The Cyprus Institute of Neurology and Genetics, 6 International Airport Avenue, Nicosia 1683, Cyprus.
E-mail: Lederer@cing.ac.cy
[4]These authors contributed equally to this work.
Received 17 October 2014; revised 1 April 2015; accepted 6 May 2015

a duplex, cross-species assay for LV VCN quantitation towards the establishment of a community standard for VCN reporting.

RESULTS AND DISCUSSION

Aiming to establish a qPCR assay with cross-species compatibility for human and murine sequences, we chose a published LV qPCR assay,[9] qLV (see Table 1), and selected the highly conserved poly (rC)-binding protein 2 (*PCBP2*) gene from the literature,[10] and a *PCBP2* region of high sequence identity and sufficient length for optimised endogene primer and probe design. The corresponding qPCR assay, qPCBP2, has perfect sequence identity of primers and probes for 35 different mammalian species (see Figure 1a), including human, apes, new- and old-world monkeys, Chinese and golden hamster, mouse, pig, domestic and wild ruminants, delphinidae, three bat species, hedgehog, shrew and armadillo, and sequence identity of the entire amplicon for 33 of these. In the following, we validate the VCN assay in human and murine cells and demonstrate the functionality of the qPCBP2 assay in three additional species, but all of the above species are either naturally or experimentally infected with lentiviruses (see, for example,[11]) and thus possible assay targets.

Duplex PCR is subject to competition of all same-tube amplicons, which for quantitative assays entails the confirmation of linear amplification for all products of interest within the prospective range and ratios of target concentrations. Following a simple workflow (see Supplementary Materials and Methods 1), we used gDNA from human and murine erythroleukaemia cell lines (HEL and MEL, respectively) to work towards a standard assay for 100 ng gDNA of lentivirally transduced sample. To this end, we tested the amplification of simplex versus duplex reactions for combinations of 50–200 ng gDNA, in order to allow for limiting sample amounts or considerable imprecision in spectrophotometrically determined DNA quantities, with lentiviral plasmid DNA (see Supplementary Materials and Methods 2). We used LV plasmid amounts equivalent to VCN 0.16 up to 20 in 100 ng of gDNA, adjusting plasmid amounts for the respective genome size and arguing that a VCN outside this range would not be used for titre calculations, and that for therapeutic applications a lower VCN would be insufficient and a higher VCN would carry an unacceptably high risk of insertional mutagenesis.

Initial attempts to establish a duplex protocol based on the kit routinely used in our laboratory for simplex detections (TaqMan 2× PCR Master Mix, Applied Biosystems) led to a depression of amplicon detection in duplex compared with equivalent simplex reactions when the alternative template became more abundant within physiological and routine experimental parameters (VCN $\geqslant$1.6 or gDNA $\geqslant$100 ng; data not shown). This nonlinear behaviour of duplex reactions persisted despite modification of a multitude of parameters relevant to qPCR efficiency and amplicon competition (see Materials and Methods), but disappeared after a switch to the 2× Multiplex PCR Master Mix (Qiagen, Hilden, Germany).

The dedicated multiplex kit in combination with qLV and qPCBP2 allowed us to established a standardised duplex protocol (see Supplementary Materials and Methods 3 and 4) with extreme correspondence between expected and observed values for LV detection (Supplementary Figure S1) and gDNA detection (Supplementary Figure S2), and between simplex and duplex detection of LV (Figure 2a and Supplementary Figure S3) and gDNA (Figure 2b) within the above parameters. The qPCR duplex assay displayed average amplification efficiencies of (96.8 ± 3.2)% for qLV and of (93.3 ± 4.8)% for qPCBP2 across all combinations of VCNs and gDNA template types and concentrations tested, as shown in Supplementary Table S1, which also gives detailed results for efficiencies and quality of fit for individual combinations.

Finally, and using gDNA from clones of MEL and HEL cells transduced with LVs typical for GT applications, we ascertained the consistency of the measured number of integrated proviral copies with those obtained through an independent simplex assay (see Figures 1b–e and Supplementary Figure S4). Subsequent tests of the assay with gDNA from additional species (see Supplementary Figure S5, demonstrating use with bovine, caprine and ovine gDNA) confirm its potential applicability to a wide range of model systems. The duplex VCN assay has since been adopted in our laboratory for the determination of LV titres after vector production and of VCN after transduction of human peripheral blood-derived $CD34^+$ cells and murine bone marrow-derived lineage-negative cells, and has proven robust and economical. Besides these benefits in its local application, the wide species compatibility of this assay allows its wider adoption and thus its establishment as a standard for comparing derived experimental and vector parameters, such as transduction efficiency and activity per vector copy, between independent lentiviral GT studies.

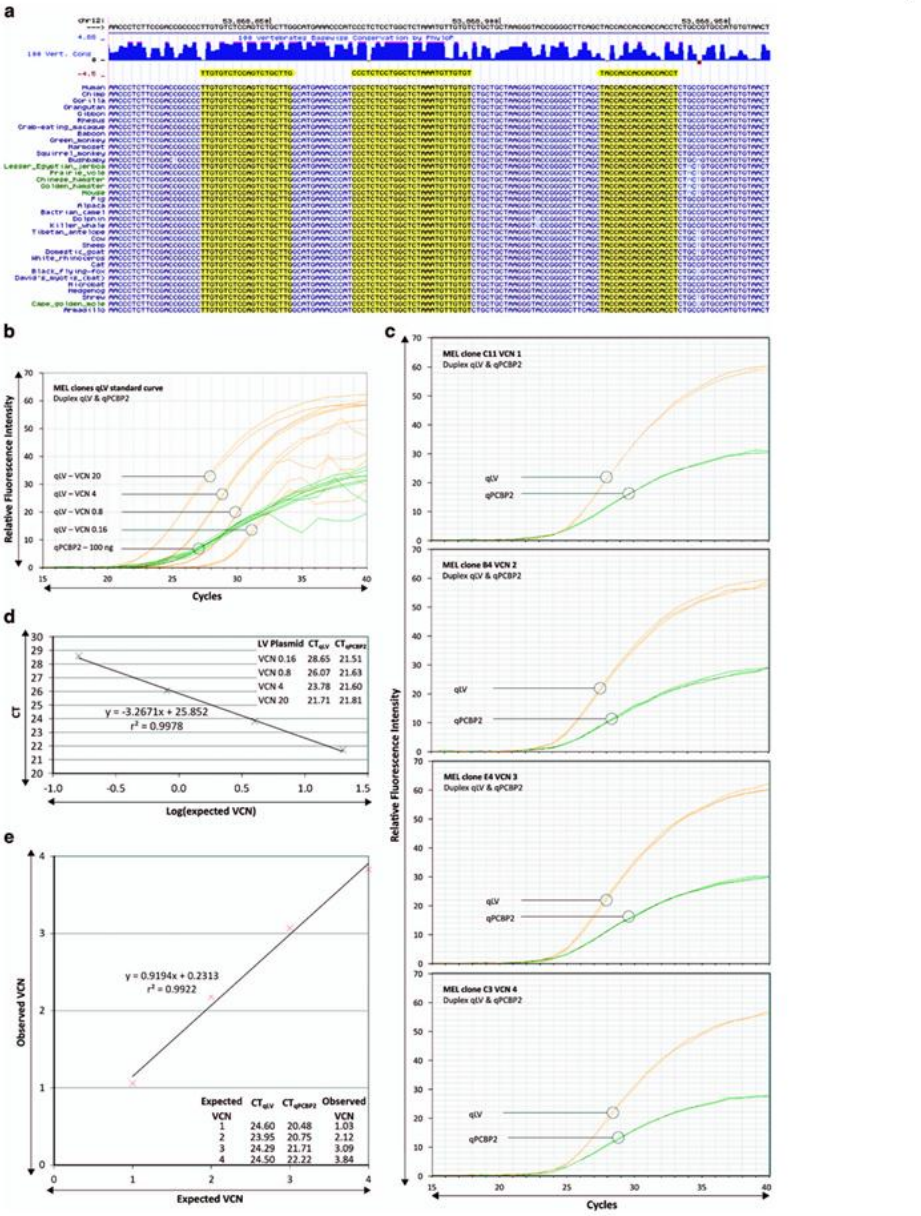
MATERIALS AND METHODS

Cell material for DNA extraction was wild-type and LV-transduced HEL (92.1.7 (ATCC TIB-180)) and MEL (48 (ATCC, Manassas, VA, USA, CRL-1913)) cell populations and clones, maintained at 37 °C, 5% CO_2 and 100% relative humidity in RPMI medium, supplemented with 1% penicillin/streptomycin, 10% fetal bovine serum and 1% L-glutamate (all from Life Technologies, Carlsbad, CA, USA). The LVs used for transduction were $MA821^{T87Q}$, a derivative of the GLOBE vector[12] encoding a T87Q mutation in the β-globin open-reading frame, and pCCL.sin.cPPT.hPGK.GFP.WPRE[13] for MEL and human erythroleukaemia cells, respectively. Cell clones were isolated by twofold limiting dilution and visual inspection of single-cell status, followed by simplex qPCR quantification using LV plasmid and gDNA standard curves. The clones used here

Table 1. Primer and probe sequences for the duplex cross-species VCN qPCR assay

qPCR assay	*Target*	*Function*	*Sequence (5′ to 3′)*
qLV	HIV gag	FW	TCTCGACGCAGGACTCG
		RV	TACTGACGCTCTCGCACC
		Probe	*Yakima-Yellow*-ATCTCTCTCCTTCTAGCCTC-*ZNA$_4$-BHQ1*
qPCBP2	*PCBP2* IVS 13	FW	TTGTGTCTCCAGTCTGCTTG
		RV	AGGTGGTGGTGGTGGTA
		Probe	FAM-CCCTCTCCTGGCTCTAAATGTTGTGT-BHQ1

Abbreviations: BHQ1, black hole quencher-1; FAM, 6-carboxyfluorescein; FW, forward primer; IVS, intervening sequence; probe, 5′ fluorescently labelled and 3′ quencher-linked probe; RV, reverse primer; ZNA, zip nucleic acid cationic spermine residues. Assay qLV is based on sequences published by Amendola *et al*.,[9] with modifications shown in italics, and recognises the gag sequence of the $MA821^{T87Q}$ transfer vector plasmid and of proviral integration sites. Assay qPCBP2 recognises an intronic *PCBP2* region of perfect cross-species identity between mouse, human and 31 additional species for the entire amplicon.



showed a consistent quantity ratio for the LV transgene in three independent experiments. gDNA was extracted using either a standard phenol:chloroform:isopropanol–chloroform extraction method (for bovine, caprine and ovine tissues) and the FlexiGene DNA kit (Qiagen, for cell lines). The LV plasmid standard used was the corresponding 9771-bp plasmid encoding the MA821^{T87Q} transfer vector segment, with 3.26 and 3.74 pg plasmid DNA representing a VCN of 1 in 100 ng of human (3.1 Gb per haploid genome, GRCh37) and murine (2.7 Gb per haploid genome, GRCm38) gDNA, respectively. Experimental procedures followed for vector production in HEK 293T cells, vector titration and lentiviral transduction of cells have been described elsewhere.[12,14] For LV standard curves, fivefold serial dilutions of 6.3 and 7.2 pg plasmid per 25-µl reaction were used in combination with human and murine gDNA, respectively. The curves thus represented VCN equivalents of 20, 4, 0.8 and 0.16 for 100 ng gDNA, with circular

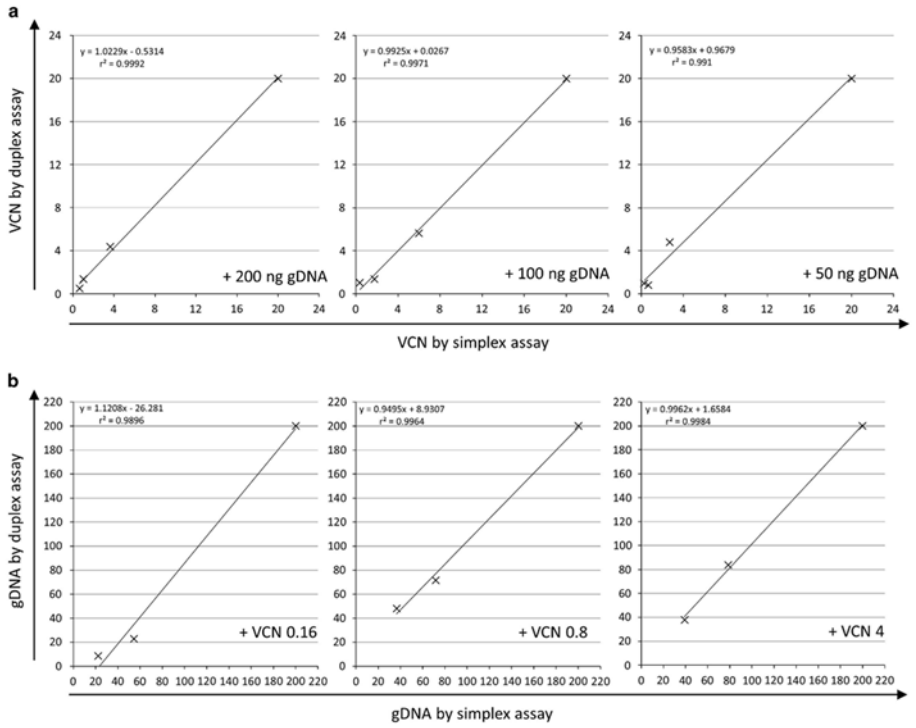


Figure 2. Simplex versus duplex quantification of LV plasmid and gDNA. The scatter plots show the correlation of template amounts determined by simplex assay with those determined by duplex assay in the presence of varying amounts of the respective alternative template, giving the formula of the trend line and the correlation (r^2) between simplex and duplex values for each plot. (**a**) Quantification of LV plasmid DNA (VCN equivalence) in the presence of varying amounts (200, 100 and 50 ng) of murine gDNA. (**b**) Quantification of murine gDNA (ng) in the presence of varying LV plasmid DNA amounts (equivalent to VCN 0.16, 0.8 and 4).

Figure 1. Sequence alignment of qPCBP2 and validation of qLV and qPCBP2 qPCR assays. (**a**) Cross-species alignment for target sequences of the *PCBP2*-specific qPCR assay *PCBP2* gene sequences were accessed using the UCSC Genome Browser (http://genome.ucsc.edu; Build GRCh37/hg19) and its Vertebrate Multiz Alignment track. Perfect sequence matches for *PCBP2* probe and primer sequences are highlighted. (**b–d**) Duplex standard curves for qLV and qPCBP2, and same-plate amplification of MEL clones with a known VCN of 1, 2, 3 and 4, respectively, for the MA821^{T87Q} LV are shown and analysed. For scatter plots, corresponding trend line formulas and curve fit values (r^2) are shown. All outlier-removed reactions of the duplex assay are shown. (**b**) Fluorescence plot of duplex qPCR for LV plasmid standard curve in the presence of 100 ng MEL gDNA, used for the absolute quantification of LV integration sites in MEL clones. The cause of plateau-phase noise is unknown. (**c**) Fluorescence plot of duplex qPCR for MEL clones with known numbers of viral integration sites. (**d**) Scatter plot of the standard curve shown in **b**. The plot shows the logarithm of expected VCNs against the automatically determined threshold cycle (CT) values and provides average CT values for each data point for qLV and qPCBP2 detection as a table insert. (**e**) Scatter plot for VCN determination of MEL clones analysed in **c**. The plot shows the expected VCN against that determined (observed) using the standard curve in **d**. The expected and observed VCN values and average CT values are additionally shown as a table insert.

and linearised plasmid (using *ScaI*-HF (New England Biolabs, Inc., Ipswich, MA, USA) for linearisation) performing equivalently. *PCBP2* was detected as an endogene in defined amounts of gDNA, both for transgenic samples and for standard curves, which comprised serial dilutions to 200, 100 and 50 ng of wild-type DNA per 25-µl reaction, either neat or laced with the MA821^{T87Q} transfer vector plasmid at the VCN equivalents given above. Quantification of viral DNA and gDNA in simplex reactions was initially performed with the TaqMan 2× PCR Master Mix (Applied Biosystems). Of the quantitative PCR assays (all from Metabion International AG, Planegg/Steinkirchen, Germany), qLV for detection of LV sequences was based on a previous publication by Amendola *et al.*[9] (see Table 1), and the endogene assay qPCBP2 recognised one of the *PCBP2* regions of perfect cross-species sequence identity long enough to allow primer and probe design, and identical amplicon composition (Vector NTI Advance 11.5, Life Technologies) (see Figure 1a). Additional assays qLV-b (FW 5′-TGAAAGCGAAAGGGAAACCA-3′, RV 5′-*TTG*CCGTGCG CGCTTCAG-3′; Probe *HEX*-5′-AGCTCTCTCGACGCAGGACTCGGC-3′-*TAMRA*, based on Miccio *et al.*[12]) and qPCBP2-b (FW 5′-CTGCAT AATCGGGCGTCAAG-3′, RV 5′-GCAGCAGATCCAGTGATGGTAACCT -3′; FAM-5′-CGCCAAAATCAATGAGATCCGTCAGATGTCT-3′-TAMRA, which inadvertently co-amplifies a human pseudogene, *PCBP2P2*, with a single mismatch (underlined) in the probe sequence) were additionally used in simplex reactions and during the optimisation of duplex assays. Attempts to establish a reliable duplex assay for the above range of gDNA and LV amounts included the independent modulation of many PCR parameters, including probe concentration (50, 100, 250 or 375 nM), primer concentration (300, 600 or 900 nM), $MgCl_2$ concentration (3–6 mM), dNTP concentration (200–400 µM), annealing temperature (60, 62 and 64 °C) and the use of alternative qPCR assays qLV-b with qPCBP2-b and qLV with qPCBP2.

All results shown in this study are based on the 2× Multiplex PCR Master Mix (Qiagen), using qPCR assays qLV and qPCBP2 in triplicate 25-µl reactions (1× Master Mix, 250 nM for each probe and 300 nM for each primer), and the 7900HT Fast Real-Time PCR System running SDS v2.4 and the standard PCR programme of 2 min 50 °C, 10 min 95 °C, 40× (15 s 95 °C, 1 min 60 °C) (all from Applied Biosystems). To aid reproduction of the analyses performed here, a simple workflow chart (Supplementary Materials and Methods 1) and Excel (Office 2010, Microsoft, Redmond, WA, USA) spread sheets for the calculation of LV plasmid and gDNA standard curves (Supplementary Materials and Methods 2), for the preparation of reactions for 96-well plates (Supplementary Materials and Methods 3) and for the postrun determination of VCN (Supplementary Materials and Methods 4), respectively, are provided as Supplementary Information. In brief, to avoid bias in the quantification of samples, all threshold cycle (CT) values were determined using automatic calculation of threshold levels in SDS v2.4. All samples were run in triplicate reactions and outliers removed automatically before statistical analyses (all Microsoft Excel), choosing a deviation of 2 s.d. from the triplicate median as a cutoff value. Statistical comparisons (all Microsoft Excel) between expected and observed VCN or gDNA values by duplex assay, and between simplex and duplex results for observed VCN or gDNA values were based on outlier-removed averages of triplicate repeats. For VCN standard curves and quantification of LV insertion sites in cellular clones, relative and absolute VCN values, respectively, were calculated using the ΔΔCT method values with *PCBP2* as the calibrator gene. For analyses of expected versus observed *PCBP2* values, no additional calibrator gene was used and relative *PCBP2* quantities were instead determined by subtracting the smallest *PCBP2* CT value *n* from all the automatically determined CT values x_i of the same run to determine the relative quantity $2^{(xi-n)}$ for each sample. (This can be emulated in Supplementary Materials and Methods 4 by filling in *PCBP2* values for the test gene (column C)] and setting all CT values for the calibrator (column H) to 0).

Amplification efficiencies and the squared Pearson's correlation coefficient r^2 of the curve fit for the log-linear phase of the qPCR amplification were calculated using the program LinRegPCR (v2015.2),[15] by exporting ΔRN values from the SDS v2.4 acquisition software, dismissing outlier-removed reactions and using at least four data points per reaction for fit calculations.

VCNs of LV clones in cell lines were isolated by twofold limiting dilution and selection of clones by visual confirmation of single-cell status after seeding. Absolute copy numbers were determined using standard curves of gDNA dilutions and LV-plasmid dilutions in simplex reactions based on qLV and qPCPB2, and, for MEL cells, additional LV and *PCBP2* qPCR assays. Clones moreover consistently showed the expected LV quantity ratios in relative qPCR comparisons in three independent experiments.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was co-funded by the European Union's Seventh Framework Program for Research, Technological Development and Demonstration under grant agreement number 306201 (THALAMOSS) and by the Republic of Cyprus through the Research Promotion Foundation under grant agreement ΥΓΕΙΑ/ΒΙΟΣ/0311(ΒΕ)/20 and through core funding of the Cyprus Institute of Neurology and Genetics.

REFERENCES

1 Pluta K, Kacprzak MM. Use of HIV as a gene transfer vector. *Acta Biochim Pol* 2009; **56**: 531–595.
2 Cavazza A, Moiani A, Mavilio F. Mechanisms of retroviral integration and mutagenesis. *Hum Gene Ther* 2013; **24**: 119–131.
3 Schuesler T, Reeves L, Kalle C, Grassman E. Copy number determination of genetically-modified hematopoietic stem cells. *Methods Mol Biol* 2009; **506**: 281–298.
4 Cooper AR, Patel S, Senadheera S, Plath K, Kohn DB, Hollis RP. Highly efficient large-scale lentiviral vector concentration by tandem tangential flow filtration. *J Virol Methods* 2011; **177**: 1–9.
5 Romero Z, Urbinati F, Geiger S, Cooper AR, Wherley J, Kaufman ML *et al.* beta-globin gene transfer to human bone marrow for sickle cell disease. *J Clin Invest* 2013.
6 Joglekar AV, Hollis RP, Kuftinec G, Senadheera S, Chan R, Kohn DB. Integrase-defective lentiviral vectors as a delivery platform for targeted modification of adenosine deaminase locus. *Mol Ther* 2013; **21**: 1705–1717.
7 Giannoni F, Hardee CL, Wherley J, Gschweng E, Senadheera S, Kaufman ML *et al.* Allelic exclusion and peripheral reconstitution by TCR transgenic T cells arising from transduced human hematopoietic stem/progenitor cells. *Mol Ther* 2013; **21**: 1044–1054.
8 Karumbayaram S, Lee P, Azghadi SF, Cooper AR, Patterson M, Kohn DB *et al.* From skin biopsy to neurons through a pluripotent intermediate under Good Manufacturing Practice protocols. *Stem Cells Transl Med* 2012; **1**: 36–43.
9 Amendola M, Passerini L, Pucci F, Gentner B, Bacchetta R, Naldini L. Regulated and multiple miRNA and siRNA delivery into primary cells by a lentiviral platform. *Mol Ther* 2009; **17**: 1039–1052.
10 Makeyev AV, Chkheidze AN, Liebhaber SA. A set of highly conserved RNA-binding proteins, alphaCP-1 and alphaCP-2, implicated in mRNA stabilization, are coexpressed from an intronless gene and its intron-containing paralog. *J Biol Chem* 1999; **274**: 24849–24857.
11 Gifford RJ. Viral evolution in deep time: lentiviruses and mammals. *Trends Genet* 2012; **28**: 89–100.
12 Miccio A, Cesari R, Lotti F, Rossi C, Sanvito F, Ponzoni M *et al.* In vivo selection of genetically modified erythroblastic progenitors leads to long-term correction of beta-thalassemia. *Proc Natl Acad Sci USA* 2008; **105**: 10547–10552.
13 Follenzi A, Ailles LE, Bakovic S, Geuna M, Naldini L. Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. *Nat Genet* 2000; **25**: 217–222.

14 Roselli EA, Mezzadra R, Frittoli MC, Maruggi G, Biral E, Mavilio F *et al*. Correction of beta-thalassemia major by gene transfer in haematopoietic progenitors of pediatric patients. *EMBO Mol Med* 2010; **2**: 315–328.
15 Ramakers C, Ruijter JM, Deprez RH, Moorman AF. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci Lett* 2003; **339**: 62–66.

Supplementary Information accompanies this paper on Gene Therapy website (http://www.nature.com/gt)

7.0 References

1. Trent RJ: **Diagnosis of the haemoglobinopathies**. *The Clinical biochemist Reviews / Australian Association of Clinical Biochemists* 2006, **27**(1):27-38.
2. Giege R: **A historical perspective on protein crystallization from 1840 to the present day**. *The FEBS journal* 2013, **280**(24):6456-6497.
3. Papadopoulos S, Jurgens KD, Gros G: **Protein diffusion in living skeletal muscle fibers: dependence on protein size, fiber type, and contraction**. *Biophysical journal* 2000, **79**(4):2084-2094.
4. Fermi G, Perutz MF, Shaanan B, Fourme R: **The crystal structure of human deoxyhaemoglobin at 1.74 A resolution**. *Journal of molecular biology* 1984, **175**(2):159-174.
5. Berg JM TJ, Stryer L.: **Hemoglobin Transports Oxygen Efficiently by Binding Oxygen Cooperatively**. In: *Biochemistry.* 5 edn. New York, U.S.A.: W H Freeman; 2002.
6. Koury S, Yarlagadda S, Moskalik-Liermo K, Popli N, Kim N, Apolito C, Peterson A, Zhang X, Zu P, Tamburlin J *et al*: **Differential gene expression during terminal erythroid differentiation**. *Genomics* 2007, **90**(5):574-582.
7. Hattangadi SM, Wong P, Zhang L, Flygare J, Lodish HF: **From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications**. *Blood* 2011, **118**(24):6258-6268.
8. Pina C, Enver T: **Differential contributions of haematopoietic stem cells to foetal and adult haematopoiesis: insights from functional analysis of transcriptional regulators**. *Oncogene* 2007, **26**(47):6750-6765.
9. Palis J: **Primitive and definitive erythropoiesis in mammals**. *Frontiers in physiology* 2014, **5**(3).
10. Becker AJ, McCulloch EA, Till JE: **Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells**. *Nature* 1963, **197**:452-454.
11. Bryder D, Rossi DJ, Weissman IL: **Hematopoietic stem cells: the paradigmatic tissue-specific stem cell**. *The American journal of pathology* 2006, **169**(2):338-346.
12. Seita J, Weissman IL: **Hematopoietic stem cell: self-renewal versus differentiation**. *Wiley interdisciplinary reviews Systems biology and medicine* 2010, **2**(6):640-653.
13. Martinez-Agosto JA, Mikkola HK, Hartenstein V, Banerjee U: **The hematopoietic stem cell and its niche: a comparative view**. *Genes & development* 2007, **21**(23):3044-3060.
14. Fuchs E: **The tortoise and the hair: slow-cycling cells in the stem cell race**. *Cell* 2009, **137**(5):811-819.
15. Reya T: **Regulation of hematopoietic stem cell self-renewal**. *Recent progress in hormone research* 2003, **58**:283-295.
16. Fibbe WE, Zijlmans JM, Willemze R: **Differential short-term and long-term repopulating ability of stem cell subsets in mice**. *Stem cells (Dayton, Ohio)* 1997, **15 Suppl 1**:47-51; discussion 52-43.
17. Orkin SH, Zon LI: **Hematopoiesis: an evolving paradigm for stem cell biology**. *Cell* 2008, **132**(4):631-644.
18. Lo Celso C, Scadden DT: **The haematopoietic stem cell niche at a glance**. *Journal of cell science* 2011, **124**(Pt 21):3529-3535.
19. Silberstein LE, Lin CP: **A new image of the hematopoietic stem cell vascular niche**. *Cell stem cell* 2013, **13**(5):514-516.
20. Chotinantakul K, Leeanansaksiri W: **Hematopoietic stem cell development, niches, and signaling pathways**. *Bone marrow research* 2012, **2012**:270425.
21. Johns JL, Christopher MM: **Extramedullary hematopoiesis: a new look at the underlying stem cell niche, theories of development, and occurrence in animals**. *Veterinary pathology* 2012, **49**(3):508-523.
22. Chasis JA, Mohandas N: **Erythroblastic islands: niches for erythropoiesis**. *Blood* 2008, **112**(3):470-478.
23. Ramos P, Casu C, Gardenghi S, Breda L, Crielaard BJ, Guy E, Marongiu MF, Gupta R, Levine RL, Abdel-Wahab O *et al*: **Macrophages support pathological erythropoiesis in polycythemia vera and beta-thalassemia**. *Nature medicine* 2013, **19**(4):437-445.
24. Chow A, Huggins M, Ahmed J, Hashimoto D, Lucas D, Kunisaki Y, Pinho S, Leboeuf M, Noizat C, van Rooijen N *et al*: **CD169(+) macrophages provide a niche promoting**

erythropoiesis under homeostasis and stress. *Nature medicine* 2013, **19**(4):429-436.
25. Elliott S, Sinclair AM: **The effect of erythropoietin on normal and neoplastic cells**. *Biologics : targets & therapy* 2012, **6**:163-189.
26. Liu J, Zhang J, Ginzburg Y, Li H, Xue F, De Franceschi L, Chasis JA, Mohandas N, An X: **Quantitative analysis of murine terminal erythroid differentiation in vivo: novel method to study normal and disordered erythropoiesis**. *Blood* 2013, **121**(8):e43-49.
27. Chen K, Liu J, Heck S, Chasis JA, An X, Mohandas N: **Resolving the distinct stages in erythroid differentiation based on dynamic changes in membrane protein expression during erythropoiesis**. *Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**(41):17413-17418.
28. Hoffman R, et al: **Biology of Erythropoiesis, Erythroid Differentiation, and Maturation**. In: *Hematology: Basic Principles and Practice.* 6 edn. PA, U.S.A.: Elsevier Saunders; 2013: 258 - 265.
29. Singh V.K. SA, et al: **Manufacturing blood ex vivo: a futuristic approach to deal with the supply and safety concerns**. *Frontiers in Cell and Developmental Biology* 2014, **2**.
30. Loffler H, Rastetter, J., Haferlach, T.: **Blood and Bone Marrow**. In: *Atlas of Clinical Hematology.* 6 edn. NY, U.S.A.: Springer-Verlag Berlin Heidelberg; 2005: 28 - 31.
31. McGrath KE, Kingsley PD, Koniski AD, Porter RL, Bushnell TP, Palis J: **Enucleation of primitive erythroid cells generates a transient population of "pyrenocytes" in the mammalian fetus**. *Blood* 2008, **111**(4):2409-2417.
32. Yoshida H, Kawane K, Koike M, Mori Y, Uchiyama Y, Nagata S: **Phosphatidylserine-dependent engulfment by macrophages of nuclei from erythroid precursor cells**. *Nature* 2005, **437**(7059):754-758.
33. Griffiths RE, Kupzig S, Cogan N, Mankelow TJ, Betin VM, Trakarnsanga K, Massey EJ, Parsons SF, Anstee DJ, Lane JD: **The ins and outs of human reticulocyte maturation: autophagy and the endosome/exosome pathway**. *Autophagy* 2012, **8**(7):1150-1151.
34. Franco RS: **Measurement of red cell lifespan and aging**. *Transfusion medicine and hemotherapy : offizielles Organ der Deutschen Gesellschaft fur Transfusionsmedizin und Immunhamatologie* 2012, **39**(5):302-307.
35. Gallagher PG: **Red cell membrane disorders**. *Hematology / the Education Program of the American Society of Hematology American Society of Hematology Education Program* 2005:13-18.
36. Franco RS, Puchulu-Campanella ME, Barber LA, Palascak MB, Joiner CH, Low PS, Cohen RM: **Changes in the properties of normal human red blood cells during in vivo aging**. *American journal of hematology* 2013, **88**(1):44-51.
37. Gifford SC, Derganc J, Shevkoplyas SS, Yoshida T, Bitensky MW: **A detailed study of time-dependent changes in human red blood cells: from reticulocyte maturation to erythrocyte senescence**. *British journal of haematology* 2006, **135**(3):395-404.
38. Smith JA: **Exercise, training and red blood cell turnover**. *Sports medicine (Auckland, NZ)* 1995, **19**(1):9-31.
39. Spivak JL: **The anaemia of cancer: death by a thousand cuts**. *Nature reviews Cancer* 2005, **5**(7):543-555.
40. Psaltis PJ, Harbuzariu A, Delacroix S, Holroyd EW, Simari RD: **Resident vascular progenitor cells--diverse origins, phenotype, and function**. *Journal of cardiovascular translational research* 2011, **4**(2):161-176.
41. P. J: **Yolk Sac Development in Mice**. In: *Hematopoietic Stem Cell Development.* Edited by Godin I CA. U.S.A.: Eurekah.com and Kluwer Academic / Plenum Publisher; 2006: 62 - 71.
42. McGrath K, Palis J: **Ontogeny of erythropoiesis in the mammalian embryo**. *Current topics in developmental biology* 2008, **82**:1-22.
43. Wang LD, Wagers AJ: **Dynamic niches in the origination and differentiation of haematopoietic stem cells**. *Nature reviews Molecular cell biology* 2011, **12**(10):643-655.
44. Palis J, Robertson S, Kennedy M, Wall C, Keller G: **Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse**. *Development (Cambridge, England)* 1999, **126**(22):5073-5084.

45. Fraser ST: **The modern primitives: applying new technological approaches to explore the biology of the earliest red blood cells**. *ISRN hematology* 2013, **2013**:568928.
46. Tanaka Y, Hayashi M, Kubota Y, Nagai H, Sheng G, Nishikawa S, Samokhvalov IM: **Early ontogenic origin of the hematopoietic stem cell lineage**. *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(12):4515-4520.
47. Medvinsky A, Rybtsov S, Taoudi S: **Embryonic origin of the adult hematopoietic system: advances and questions**. *Development (Cambridge, England)* 2011, **138**(6):1017-1031.
48. de Bruijn MF, Speck NA, Peeters MC, Dzierzak E: **Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo**. *Embo J* 2000, **19**(11):2465-2474.
49. Ivanovs A, Rybtsov S, Welch L, Anderson RA, Turner ML, Medvinsky A: **Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region**. *The Journal of experimental medicine* 2011, **208**(12):2417-2427.
50. Moignard V, Woodhouse S, Fisher J, Gottgens B: **Transcriptional hierarchies regulating early blood cell development**. *Blood cells, molecules & diseases* 2013, **51**(4):239-247.
51. Weatherall DJ, Clegg JB: **Inherited haemoglobin disorders: an increasing global health problem**. *Bulletin of the World Health Organization* 2001, **79**(8):704-712.
52. Kohne E: **Hemoglobinopathies: clinical manifestations, diagnosis, and treatment**. *Deutsches Arzteblatt international* 2011, **108**(31-32):532-540.
53. Olivieri NF: **The beta-thalassemias**. *The New England journal of medicine* 1999, **341**(2):99-109.
54. Weatherall D: **2003 William Allan Award address. The Thalassemias: the role of molecular genetics in an evolving global health problem**. *American journal of human genetics* 2004, **74**(3):385-392.
55. Higgs DR, Engel JD, Stamatoyannopoulos G: **Thalassaemia**. *Lancet* 2012, **379**(9813):373-383.
56. Birgens H, Ljung R: **The thalassaemia syndromes**. *Scandinavian journal of clinical and laboratory investigation* 2007, **67**(1):11-25.
57. Thein SL: **The molecular basis of beta-thalassemia**. *Cold Spring Harbor perspectives in medicine* 2013, **3**(5):a011700.
58. Cao A, Moi P, Galanello R: **Recent advances in beta-thalassemias**. *Pediatric reports* 2011, **3**(2):e17.
59. Thein SL, Menzel S, Lathrop M, Garner C: **Control of fetal hemoglobin: new insights emerging from genomics and clinical implications**. *Human molecular genetics* 2009, **18**(R2):R216-223.
60. Thein SL, Menzel S: **Discovering the genetics underlying foetal haemoglobin production in adults**. *British journal of haematology* 2009, **145**(4):455-467.
61. Borg J, Papadopoulos P, Georgitsi M, Gutierrez L, Grech G, Fanis P, Phylactides M, Verkerk AJ, van der Spek PJ, Scerri CA *et al*: **Haploinsufficiency for the erythroid transcription factor KLF1 causes hereditary persistence of fetal hemoglobin**. *Nature genetics* 2010, **42**(9):801-805.
62. Jawaid K, Wahlberg K, Thein SL, Best S: **Binding patterns of BCL11A in the globin and GATA1 loci and characterization of the BCL11A fetal hemoglobin locus**. *Blood cells, molecules & diseases* 2010, **45**(2):140-146.
63. Xu J, Sankaran VG, Ni M, Menne TF, Puram RV, Kim W, Orkin SH: **Transcriptional silencing of {gamma}-globin by BCL11A involves long-range interactions and cooperation with SOX6**. *Genes & development* 2010, **24**(8):783-798.
64. Rachmilewitz EA, Giardina PJ: **How I treat thalassemia**. *Blood* 2011, **118**(13):3479-3488.
65. Modell B, Darlison M: **Global epidemiology of haemoglobin disorders and derived service indicators**. *Bulletin of the World Health Organization* 2008, **86**(6):480-487.
66. Angastiniotis MA, Hadjiminas MG: **Prevention of thalassaemia in Cyprus**. *Lancet* 1981, **1**(8216):369-371.

67. Angastiniotis M, Modell B: **Global epidemiology of hemoglobin disorders**. *Annals of the New York Academy of Sciences* 1998, **850**:251-269.
68. Cao A: **1993 William Allan award address**. *American journal of human genetics* 1994, **54**(3):397-402.
69. Unit HTA: **Management of Thalassaemia**. In. Edited by Division MD. Ministry of Health Malaysia.
70. Rund D, Rachmilewitz E: **Beta-thalassemia**. *The New England journal of medicine* 2005, **353**(11):1135-1146.
71. Adam A: **Review of oral iron chelators for the treatment of iron overload in pediatric patients**. In.
72. Olivieri NF, Brittenham GM: **Iron-chelating therapy and the treatment of thalassemia**. *Blood* 1997, **89**(3):739-761.
73. Kwiatkowski JL: **Oral iron chelators**. *Pediatric clinics of North America* 2008, **55**(2):461-482, x.
74. Thomas ED, Buckner CD, Sanders JE, Papayannopoulou T, Borgna-Pignatti C, De Stefano P, Sullivan KM, Clift RA, Storb R: **Marrow transplantation for thalassaemia**. *Lancet* 1982, **2**(8292):227-229.
75. Sadelain M, Boulad F, Galanello R, Giardina P, Locatelli F, Maggio A, Rivella S, Riviere I, Tisdale J: **Therapeutic options for patients with severe beta-thalassemia: the need for globin gene therapy**. *Human gene therapy* 2007, **18**(1):1-9.
76. La Nasa G, Giardini C, Argiolu F, Locatelli F, Arras M, De Stefano P, Ledda A, Pizzati A, Sanna MA, Vacca A *et al*: **Unrelated donor bone marrow transplantation for thalassemia: the effect of extended haplotypes**. *Blood* 2002, **99**(12):4350-4356.
77. Chandrakasan S, Malik P: **Gene therapy for hemoglobinopathies: the state of the field and the future**. *Hematology/oncology clinics of North America* 2014, **28**(2):199-216.
78. Walters MC: **Gene therapy and bone marrow transplantation for thalassemia: changing of the guard?** *Mol Ther* 2010, **18**(9):1577.
79. Higgs DR, Vickers MA, Wilkie AO, Pretorius IM, Jarman AP, Weatherall DJ: **A review of the molecular genetics of the human alpha-globin gene cluster**. *Blood* 1989, **73**(5):1081-1104.
80. Vernimmen D: **Uncovering enhancer functions using the alpha-globin locus**. *PLoS genetics* 2014, **10**(10):e1004668.
81. Li Q, Harju S, Peterson KR: **Locus control regions: coming of age at a decade plus**. *Trends in genetics : TIG* 1999, **15**(10):403-408.
82. Forget BG HR: **The normal structure and regulation of globin gene clusters**. In: *Disorders of Hemoglobin - Genetics, Pathophysiology, and Clinical Management.* Edited by Steinberg MH FB, Higgs DR, Weatherall DJ, 2 edn. NY, U.S.A.: Cambridge University Press; 2009: 46-61.
83. Schechter AN: **Hemoglobin research and the origins of molecular medicine**. *Blood* 2008, **112**(10):3927-3938.
84. Sankaran VG, Orkin SH: **The switch from fetal to adult hemoglobin**. *Cold Spring Harbor perspectives in medicine* 2013, **3**(1):a011643.
85. Albitar M, Katsumata M, Liebhaber SA: **Human alpha-globin genes demonstrate autonomous developmental regulation in transgenic mice**. *Molecular and cellular biology* 1991, **11**(7):3786-3794.
86. Schroeder WA HT, Brown AK, Uy R, Bouver NG, Lerch PO, Shelton JR, Shelton JB, Apell G: **Postnatal changes in the chemical heterogeneity of human fetal hemoglobin**. *Pediat Res* 1971, **5**:493-499.
87. Schroeder WA, Huisman TH, Shelton JR, Shelton JB, Kleihauer EF, Dozy AM, Robberson B: **Evidence for multiple structural genes for the gamma chain of human fetal hemoglobin**. *Proceedings of the National Academy of Sciences of the United States of America* 1968, **60**(2):537-544.
88. Dzierzak E, Philipsen S: **Erythropoiesis: development and differentiation**. *Cold Spring Harbor perspectives in medicine* 2013, **3**(4):a011601.
89. Weatherall DJ: **Phenotype-genotype relationships in monogenic disease: lessons from the thalassaemias**. *Nature reviews Genetics* 2001, **2**(4):245-255.

90. Sharpe JA, Chan-Thomas PS, Lida J, Ayyub H, Wood WG, Higgs DR: **Analysis of the human alpha globin upstream regulatory element (HS-40) in transgenic mice**. *Embo J* 1992, **11**(12):4565-4572.
91. Zhang HB, Liu DP, Liang CC: **The control of expression of the alpha-globin gene cluster**. *International journal of hematology* 2002, **76**(5):420-426.
92. Gourdon G, Sharpe JA, Higgs DR, Wood WG: **The mouse alpha-globin locus regulatory element**. *Blood* 1995, **86**(2):766-775.
93. Chen H, Lowrey CH, Stamatoyannopoulos G: **Analysis of enhancer function of the HS-40 core sequence of the human alpha-globin cluster**. *Nucleic acids research* 1997, **25**(14):2917-2922.
94. Levings PP, Bungert J: **The human beta-globin locus control region**. *European journal of biochemistry / FEBS* 2002, **269**(6):1589-1599.
95. Li Q, Peterson KR, Fang X, Stamatoyannopoulos G: **Locus control regions**. *Blood* 2002, **100**(9):3077-3086.
96. Bungert J, Tanimoto K, Patel S, Liu Q, Fear M, Engel JD: **Hypersensitive site 2 specifies a unique function within the human beta-globin locus control region to stimulate globin gene transcription**. *Molecular and cellular biology* 1999, **19**(4):3062-3072.
97. Fraser P, Pruzina S, Antoniou M, Grosveld F: **Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes**. *Genes & development* 1993, **7**(1):106-113.
98. Bender MA, Bulger M, Close J, Groudine M: **Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region**. *Molecular cell* 2000, **5**(2):387-393.
99. Noordermeer D, de Laat W: **Joining the loops: beta-globin gene regulation**. *IUBMB life* 2008, **60**(12):824-833.
100. Schubeler D, Groudine M, Bender MA: **The murine beta-globin locus control region regulates the rate of transcription but not the hyperacetylation of histones at the active genes**. *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**(20):11432-11437.
101. Schubeler D, Francastel C, Cimbora DM, Reik A, Martin DI, Groudine M: **Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus**. *Genes & development* 2000, **14**(8):940-950.
102. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W: **The beta-globin nuclear compartment in development and erythroid differentiation**. *Nature genetics* 2003, **35**(2):190-194.
103. Kim A, Dean A: **Chromatin loop formation in the beta-globin locus and its role in globin gene transcription**. *Molecules and cells* 2012, **34**(1):1-5.
104. Hanscombe O, Whyatt D, Fraser P, Yannoutsos N, Greaves D, Dillon N, Grosveld F: **Importance of globin gene order for correct developmental expression**. *Genes & development* 1991, **5**(8):1387-1394.
105. Smith E, Shilatifard A: **Enhancer biology and enhanceropathies**. *Nature structural & molecular biology* 2014, **21**(3):210-219.
106. Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA: **Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor**. *Cell* 2012, **149**(6):1233-1244.
107. Patrinos GP, de Krom M, de Boer E, Langeveld A, Imam AM, Strouboulis J, de Laat W, Grosveld FG: **Multiple interactions between regulatory regions are required to stabilize an active chromatin hub**. *Genes & development* 2004, **18**(12):1495-1509.
108. Dean A: **On a chromosome far, far away: LCRs and gene expression**. *Trends in genetics : TIG* 2006, **22**(1):38-45.
109. Sankaran VG, Xu J, Ragoczy T, Ippolito GC, Walkley CR, Maika SD, Fujiwara Y, Ito M, Groudine M, Bender MA *et al*: **Developmental and species-divergent globin switching are driven by BCL11A**. *Nature* 2009, **460**(7259):1093-1097.
110. Kikuchi K, Kondo M: **Developmental switch of mouse hematopoietic stem cells from fetal to adult type occurs in bone marrow after birth**. *Proceedings of the National Academy of Sciences of the United States of America* 2006, **103**(47):17852-17857.

111. Wilber A, Nienhuis AW, Persons DA: **Transcriptional regulation of fetal to adult hemoglobin switching: new therapeutic opportunities**. *Blood* 2011, **117**(15):3945-3953.
112. Nakamura T, Yamazaki Y, Saiki Y, Moriyama M, Largaespada DA, Jenkins NA, Copeland NG: **Evi9 encodes a novel zinc finger protein that physically interacts with BCL6, a known human B-cell proto-oncogene product**. *Molecular and cellular biology* 2000, **20**(9):3178-3186.
113. Liu P, Keller JR, Ortiz M, Tessarollo L, Rachel RA, Nakamura T, Jenkins NA, Copeland NG: **Bcl11a is essential for normal lymphoid development**. *Nature immunology* 2003, **4**(6):525-532.
114. Uda M, Galanello R, Sanna S, Lettre G, Sankaran VG, Chen W, Usala G, Busonero F, Maschio A, Albai G *et al*: **Genome-wide association study shows BCL11A associated with persistent fetal hemoglobin and amelioration of the phenotype of beta-thalassemia**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(5):1620-1625.
115. Lettre G, Sankaran VG, Bezerra MA, Araujo AS, Uda M, Sanna S, Cao A, Schlessinger D, Costa FF, Hirschhorn JN *et al*: **DNA polymorphisms at the BCL11A, HBS1L-MYB, and beta-globin loci associate with fetal hemoglobin levels and pain crises in sickle cell disease**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(33):11869-11874.
116. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, Hirschhorn JN: **Genome-wide association studies for complex traits: consensus, uncertainty and challenges**. *Nature reviews Genetics* 2008, **9**(5):356-369.
117. Frazer KA, Murray SS, Schork NJ, Topol EJ: **Human genetic variation and its contribution to complex traits**. *Nature reviews Genetics* 2009, **10**(4):241-251.
118. Sankaran VG, Xu J, Orkin SH: **Transcriptional silencing of fetal hemoglobin by BCL11A**. *Annals of the New York Academy of Sciences* 2010, **1202**:64-68.
119. Sankaran VG, Menne TF, Xu J, Akie TE, Lettre G, Van Handel B, Mikkola HK, Hirschhorn JN, Cantor AB, Orkin SH: **Human fetal hemoglobin expression is regulated by the developmental stage-specific repressor BCL11A**. *Science (New York, NY)* 2008, **322**(5909):1839-1842.
120. Liu H, Ippolito GC, Wall JK, Niu T, Probst L, Lee BS, Pulford K, Banham AH, Stockwin L, Shaffer AL *et al*: **Functional studies of BCL11A: characterization of the conserved BCL11A-XL splice variant and its interaction with BCL6 in nuclear paraspeckles of germinal center B cells**. *Molecular cancer* 2006, **5**:18.
121. Xu J, Bauer DE, Kerenyi MA, Vo TD, Hou S, Hsu YJ, Yao H, Trowbridge JJ, Mandel G, Orkin SH: **Corepressor-dependent silencing of fetal hemoglobin expression by BCL11A**. *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**(16):6518-6523.
122. Sankaran VG, Xu J, Orkin SH: **Advances in the understanding of haemoglobin switching**. *British journal of haematology* 2010, **149**(2):181-194.
123. Bauer DE, Kamran SC, Lessard S, Xu J, Fujiwara Y, Lin C, Shao Z, Canver MC, Smith EC, Pinello L *et al*: **An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level**. *Science (New York, NY)* 2013, **342**(6155):253-257.
124. Lefebvre V: **The SoxD transcription factors--Sox5, Sox6, and Sox13--are key cell fate modulators**. *The international journal of biochemistry & cell biology* 2010, **42**(3):429-432.
125. Lefebvre V, Dumitriu B, Penzo-Mendez A, Han Y, Pallavi B: **Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors**. *The international journal of biochemistry & cell biology* 2007, **39**(12):2195-2214.
126. Hagiwara N, Klewer SE, Samson RA, Erickson DT, Lyon MF, Brilliant MH: **Sox6 is a candidate gene for p100H myopathy, heart block, and sudden neonatal death**. *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(8):4180-4185.
127. Suzuki M, Yamamoto M, Engel JD: **Fetal globin gene repressors as drug targets for molecular therapies to treat the beta-globinopathies**. *Molecular and cellular biology* 2014.

128. Yi Z, Cohen-Barak O, Hagiwara N, Kingsley PD, Fuchs DA, Erickson DT, Epner EM, Palis J, Brilliant MH: **Sox6 directly silences epsilon globin expression in definitive erythropoiesis**. *PLoS genetics* 2006, **2**(2):e14.
129. Dumitriu B, Bhattaram P, Dy P, Huang Y, Quayum N, Jensen J, Lefebvre V: **Sox6 is necessary for efficient erythropoiesis in adult mice under physiological and anemia-induced stress conditions**. *PloS one* 2010, **5**(8):e12088.
130. Cantu C, Ierardi R, Alborelli I, Fugazza C, Cassinelli L, Piconese S, Bose F, Ottolenghi S, Ferrari G, Ronchi A: **Sox6 enhances erythroid differentiation in human erythroid progenitors**. *Blood* 2011, **117**(13):3669-3679.
131. Sripichai O, Kiefer CM, Bhanu NV, Tanno T, Noh SJ, Goh SH, Russell JE, Rognerud CL, Ou CN, Oneal PA *et al*: **Cytokine-mediated increases in fetal hemoglobin are associated with globin gene histone modification and transcription factor reprogramming**. *Blood* 2009, **114**(11):2299-2306.
132. Sankaran VG, Menne J, Heller R: **Heterozygous disruption of human SOX6 is insufficient to impair erythropoiesis or silencing of fetal hemoglobin**. *Blood* 2011, **117**(16):4396-4397.
133. Miller IJ, Bieker JJ: **A novel, erythroid cell-specific murine transcription factor that binds to the CACCC element and is related to the Kruppel family of nuclear proteins**. *Molecular and cellular biology* 1993, **13**(5):2776-2786.
134. Nuez B, Michalovich D, Bygrave A, Ploemacher R, Grosveld F: **Defective haematopoiesis in fetal liver resulting from inactivation of the EKLF gene**. *Nature* 1995, **375**(6529):316-318.
135. Perkins AC, Sharpe AH, Orkin SH: **Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF**. *Nature* 1995, **375**(6529):318-322.
136. Wijgerde M, Gribnau J, Trimborn T, Nuez B, Philipsen S, Grosveld F, Fraser P: **The role of EKLF in human beta-globin gene competition**. *Genes & development* 1996, **10**(22):2894-2902.
137. Tallack MR, Magor GW, Dartigues B, Sun L, Huang S, Fittock JM, Fry SV, Glazov EA, Bailey TL, Perkins AC: **Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq**. *Genome research* 2012, **22**(12):2385-2398.
138. Yien YY, Bieker JJ: **EKLF/KLF1, a tissue-restricted integrator of transcriptional control, chromatin remodeling, and lineage determination**. *Molecular and cellular biology* 2013, **33**(1):4-13.
139. Gallienne AE, Dreau HM, Schuh A, Old JM, Henderson S: **Ten novel mutations in the erythroid transcription factor KLF1 gene associated with increased fetal hemoglobin levels in adults**. *Haematologica* 2012, **97**(3):340-343.
140. Satta S, Perseu L, Moi P, Asunis I, Cabriolu A, Maccioni L, Demartis FR, Manunza L, Cao A, Galanello R: **Compound heterozygosity for KLF1 mutations associated with remarkable increase of fetal hemoglobin and red cell protoporphyrin**. *Haematologica* 2011, **96**(5):767-770.
141. Radmilovic M, Zukic B, Petrovic MS, Bartsakoulia M, Stankovic B, Kotur N, Dokmanovic L, Georgitsi M, Patrinos GP, Pavlovic S: **Functional analysis of a novel KLF1 gene promoter variation associated with hereditary persistence of fetal hemoglobin**. *Annals of hematology* 2013, **92**(1):53-58.
142. Zhou D, Liu K, Sun CW, Pawlik KM, Townes TM: **KLF1 regulates BCL11A expression and gamma- to beta-globin gene switching**. *Nature genetics* 2010, **42**(9):742-744.
143. Esteghamat F, Gillemans N, Bilic I, van den Akker E, Cantu I, van Gent T, Klingmuller U, van Lom K, von Lindern M, Grosveld F *et al*: **Erythropoiesis and globin switching in compound Klf1::Bcl11a mutant mice**. *Blood* 2013, **121**(13):2553-2562.
144. Stadhouders R, Aktuna S, Thongjuea S, Aghajanirefah A, Pourfarzad F, van Ijcken W, Lenhard B, Rooks H, Best S, Menzel S *et al*: **HBS1L-MYB intergenic variants modulate fetal hemoglobin via long-range MYB enhancers**. *The Journal of clinical investigation* 2014, **124**(4):1699-1710.
145. Roosjen M, McColl B, Kao B, Gearing LJ, Blewitt ME, Vadolas J: **Transcriptional regulators Myb and BCL11A interplay with DNA methyltransferase 1 in developmental silencing of embryonic and fetal beta-like globin genes**. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2014, **28**(4):1610-1620.

146. Mabaera R, West RJ, Conine SJ, Macari ER, Boyd CD, Engman CA, Lowrey CH: **A cell stress signaling model of fetal hemoglobin induction: what doesn't kill red blood cells may make them stronger**. *Experimental hematology* 2008, **36**(9):1057-1072.
147. Golan EA TA, Armstrong EJ, Armstrong AW: **Pharmacology of hematopoiesis and immunomodulation**. In: *Principles of pharmacology: The pathophysiologic basis of drug therapy.* Edited by S. R, 3 edn. PA, U.S.A.: Lippincott Williams & Wilkins; 2002: 782-784.
148. Testa U: **Fetal hemoglobin chemical inducers for treatment of hemoglobinopathies**. *Annals of hematology* 2009, **88**(6):505-528.
149. Halsey C, Roberts IA: **The role of hydroxyurea in sickle cell disease**. *British journal of haematology* 2003, **120**(2):177-186.
150. Fathallah H, Atweh GF: **Induction of fetal hemoglobin in the treatment of sickle cell disease**. *Hematology / the Education Program of the American Society of Hematology American Society of Hematology Education Program* 2006:58-62.
151. Kay MA: **State-of-the-art gene-based therapies: the road ahead**. *Nature reviews Genetics* 2011, **12**(5):316-328.
152. Fischer A, Cavazzana-Calvo M: **Gene therapy of inherited diseases**. *Lancet* 2008, **371**(9629):2044-2047.
153. O'Connor TP, Crystal RG: **Genetic medicines: treatment strategies for hereditary disorders**. *Nature reviews Genetics* 2006, **7**(4):261-276.
154. Pelletier R, Caron SO, Puymirat J: **RNA based gene therapy for dominantly inherited diseases**. *Current gene therapy* 2006, **6**(1):131-146.
155. Gaj T, Gersbach CA, Barbas CF, 3rd: **ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering**. *Trends in biotechnology* 2013, **31**(7):397-405.
156. Miccio A, Cesari R, Lotti F, Rossi C, Sanvito F, Ponzoni M, Routledge SJ, Chow CM, Antoniou MN, Ferrari G: **In vivo selection of genetically modified erythroblastic progenitors leads to long-term correction of beta-thalassemia**. *Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**(30):10547-10552.
157. Niidome T, Huang L: **Gene therapy progress and prospects: nonviral vectors**. *Gene therapy* 2002, **9**(24):1647-1652.
158. Yin H, Kanasty RL, Eltoukhy AA, Vegas AJ, Dorkin JR, Anderson DG: **Non-viral vectors for gene-based therapy**. *Nature reviews Genetics* 2014, **15**(8):541-555.
159. Mali S: **Delivery systems for gene therapy**. *Indian journal of human genetics* 2013, **19**(1):3-8.
160. Ditto AJ, Shah PN, Yun YH: **Non-viral gene delivery using nanoparticles**. *Expert opinion on drug delivery* 2009, **6**(11):1149-1160.
161. Tros de Ilarduya C, Sun Y, Duzgunes N: **Gene delivery by lipoplexes and polyplexes**. *European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences* 2010, **40**(3):159-170.
162. Yannaki E, Emery DW, Stamatoyannopoulos G: **Gene therapy for beta-thalassaemia: the continuing challenge**. *Expert reviews in molecular medicine* 2010, **12**:e31.
163. Naldini L: **Ex vivo gene transfer and correction for cell-based therapies**. *Nature reviews Genetics* 2011, **12**(5):301-315.
164. Young LS, Searle PF, Onion D, Mautner V: **Viral gene therapy strategies: from basic science to clinical application**. *The Journal of pathology* 2006, **208**(2):299-318.
165. Thomas CE, Ehrhardt A, Kay MA: **Progress and problems with the use of viral vectors for gene therapy**. *Nature reviews Genetics* 2003, **4**(5):346-358.
166. Hacein-Bey-Abina S, Le Deist F, Carlier F, Bouneaud C, Hue C, De Villartay JP, Thrasher AJ, Wulffraat N, Sorensen R, Dupuis-Girod S *et al*: **Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy**. *The New England journal of medicine* 2002, **346**(16):1185-1193.
167. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E *et al*: **LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1**. *Science (New York, NY)* 2003, **302**(5644):415-419.
168. Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, Koehl U, Glimm H, Kuhlcke K, Schilz A, Kunkel H *et al*: **Correction of X-linked chronic granulomatous disease by**

gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nature medicine* 2006, **12**(4):401-409.
169. Stein S, Ott MG, Schultze-Strasser S, Jauch A, Burwinkel B, Kinner A, Schmidt M, Kramer A, Schwable J, Glimm H *et al*: **Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease**. *Nature medicine* 2010, **16**(2):198-204.
170. Mann R, Mulligan RC, Baltimore D: **Construction of a retrovirus packaging mutant and its use to produce helper-free defective retrovirus**. *Cell* 1983, **33**(1):153-159.
171. Moiani A, Paleari Y, Sartori D, Mezzadra R, Miccio A, Cattoglio C, Cocchiarella F, Lidonnici MR, Ferrari G, Mavilio F: **Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts**. *The Journal of clinical investigation* 2012, **122**(5):1653-1666.
172. Edry E LR, Wagner S, and Rosenblum K: **Virally mediated gene manipulation in the adult CNS**. *Frontiers in Molecular Neuroscience* 2011.
173. Malik P, Arumugam PI: **Gene Therapy for beta-thalassemia**. *Hematology Am Soc Hematol Educ Program* 2005:45-50.
174. Arumugam P, Malik P: **Genetic therapy for beta-thalassemia: from the bench to the bedside**. *Hematology / the Education Program of the American Society of Hematology American Society of Hematology Education Program* 2010, **2010**:445-450.
175. Dong A, Rivella S, Breda L: **Gene therapy for hemoglobinopathies: progress and challenges**. *Translational research : the journal of laboratory and clinical medicine* 2013, **161**(4):293-306.
176. Arumugam PI, Urbinati F, Velu CS, Higashimoto T, Grimes HL, Malik P: **The 3' region of the chicken hypersensitive site-4 insulator has properties similar to its core and is required for full insulator activity**. *PloS one* 2009, **4**(9):e6995.
177. Zhang F, Thornhill SI, Howe SJ, Ulaganathan M, Schambach A, Sinclair J, Kinnon C, Gaspar HB, Antoniou M, Thrasher AJ: **Lentiviral vectors containing an enhancer-less ubiquitously acting chromatin opening element (UCOE) provide highly reproducible and stable transgene expression in hematopoietic cells**. *Blood* 2007, **110**(5):1448-1457.
178. Puthenveetil G, Scholes J, Carbonell D, Qureshi N, Xia P, Zeng L, Li S, Yu Y, Hiti AL, Yee JK *et al*: **Successful correction of the human beta-thalassemia major phenotype using a lentiviral vector**. *Blood* 2004, **104**(12):3445-3453.
179. Groth AC, Liu M, Wang H, Lovelett E, Emery DW: **Identification and characterization of enhancer-blocking insulators to reduce retroviral vector genotoxicity**. *PloS one* 2013, **8**(10):e76528.
180. Breda L, Casu C, Gardenghi S, Bianchi N, Cartegni L, Narla M, Yazdanbakhsh K, Musso M, Manwani D, Little J *et al*: **Therapeutic hemoglobin levels after gene transfer in beta-thalassemia mice and in hematopoietic cells of beta-thalassemia and sickle cells disease patients**. *PloS one* 2012, **7**(3):e32345.
181. Papapetrou EP, Lee G, Malani N, Setty M, Riviere I, Tirunagari LM, Kadota K, Roth SL, Giardina P, Viale A *et al*: **Genomic safe harbors permit high beta-globin transgene expression in thalassemia induced pluripotent stem cells**. *Nature biotechnology* 2011, **29**(1):73-78.
182. May C, Rivella S, Callegari J, Heller G, Gaensler KM, Luzzatto L, Sadelain M: **Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin**. *Nature* 2000, **406**(6791):82-86.
183. May C, Rivella S, Chadburn A, Sadelain M: **Successful treatment of murine beta-thalassemia intermedia by transfer of the human beta-globin gene**. *Blood* 2002, **99**(6):1902-1908.
184. Rivella S, May C, Chadburn A, Riviere I, Sadelain M: **A novel murine model of Cooley anemia and its rescue by lentiviral-mediated human beta-globin gene transfer**. *Blood* 2003, **101**(8):2932-2939.
185. Imren S, Payen E, Westerman KA, Pawliuk R, Fabry ME, Eaves CJ, Cavilla B, Wadsworth LD, Beuzard Y, Bouhassira EE *et al*: **Permanent and panerythroid correction of murine beta thalassemia by multiple lentiviral integration in hematopoietic stem cells**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(22):14380-14385.

186. Pawliuk R, Westerman KA, Fabry ME, Payen E, Tighe R, Bouhassira EE, Acharya SA, Ellis J, London IM, Eaves CJ *et al*: **Correction of sickle cell disease in transgenic mouse models by gene therapy**. *Science (New York, NY)* 2001, **294**(5550):2368-2371.
187. Arumugam PI, Scholes J, Perelman N, Xia P, Yee JK, Malik P: **Improved human beta-globin expression from self-inactivating lentiviral vectors carrying the chicken hypersensitive site-4 (cHS4) insulator element**. *Mol Ther* 2007, **15**(10):1863-1871.
188. Roselli EA, Mezzadra R, Frittoli MC, Maruggi G, Biral E, Mavilio F, Mastropietro F, Amato A, Tonon G, Refaldi C *et al*: **Correction of beta-thalassemia major by gene transfer in haematopoietic progenitors of pediatric patients**. *EMBO molecular medicine* 2010, **2**(8):315-328.
189. Miccio A, Poletti V, Tiboni F, Rossi C, Antonelli A, Mavilio F, Ferrari G: **The GATA1-HS2 enhancer allows persistent and position-independent expression of a beta-globin transgene**. *PloS one* 2011, **6**(12):e27955.
190. Deng W, Rupon JW, Krivega I, Breda L, Motta I, Jahn KS, Reik A, Gregory PD, Rivella S, Dean A *et al*: **Reactivation of developmentally silenced globin genes by forced chromatin looping**. *Cell* 2014, **158**(4):849-860.
191. Drakopoulou E, Papanikolaou E, Georgomanoli M, Anagnou NP: **Towards more successful gene therapy clinical trials for beta-thalassemia**. *Current molecular medicine* 2013, **13**(8):1314-1330.
192. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC: **Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans**. *Nature* 1998, **391**(6669):806-811.
193. Voinnet O: **Non-cell autonomous RNA silencing**. *FEBS letters* 2005, **579**(26):5858-5871.
194. Castel SE, Martienssen RA: **RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond**. *Nature reviews Genetics* 2013, **14**(2):100-112.
195. Ghildiyal M, Zamore PD: **Small silencing RNAs: an expanding universe**. *Nature reviews Genetics* 2009, **10**(2):94-108.
196. Carthew RW, Sontheimer EJ: **Origins and Mechanisms of miRNAs and siRNAs**. *Cell* 2009, **136**(4):642-655.
197. Pushparaj PN, Aarthi JJ, Manikandan J, Kumar SD: **siRNA, miRNA, and shRNA: in vivo applications**. *Journal of dental research* 2008, **87**(11):992-1003.
198. Hamilton AJ, Baulcombe DC: **A species of small antisense RNA in posttranscriptional gene silencing in plants**. *Science (New York, NY)* 1999, **286**(5441):950-952.
199. Zamore PD, Tuschl T, Sharp PA, Bartel DP: **RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals**. *Cell* 2000, **101**(1):25-33.
200. Tomari Y, Zamore PD: **Perspective: machines for RNAi**. *Genes & development* 2005, **19**(5):517-529.
201. Obbard. DJ. GK, . et al: **The evolution of RNAi as a defense against viruses and transposable elements**. *Philosophical Transactions of the Royal Society B* 2009, **364**:99-115.
202. Lin SL, Miller JD, Ying SY: **Intronic microRNA (miRNA)**. *J Biomed Biotechnol* 2006, **4**:26818.
203. Neilson JR, Sharp PA: **Small RNA regulators of gene expression**. *Cell* 2008, **134**(6):899-902.
204. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function**. *Cell* 2004, **116**(2):281-297.
205. Manjunath N, Wu H, Subramanya S, Shankar P: **Lentiviral delivery of short hairpin RNAs**. *Adv Drug Deliv Rev* 2009, **61**(9):732-745.
206. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC: **The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila**. *Cell* 2007, **130**(1):89-100.
207. Krol J, Sobczak K, Wilczynska U, Drath M, Jasinska A, Kaczynska D, Krzyzosiak WJ: **Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design**. *The Journal of biological chemistry* 2004, **279**(40):42230-42239.

208. Liu YP, Berkhout B: **miRNA cassettes in viral vectors: problems and solutions**. *Biochimica et biophysica acta* 2011, **1809**(11-12):732-745.
209. Gu S, Jin L, Zhang Y, Huang Y, Zhang F, Valdmanis PN, Kay MA: **The loop position of shRNAs and pre-miRNAs is critical for the accuracy of dicer processing in vivo**. *Cell* 2012, **151**(4):900-911.
210. Johanson TM, Lew AM, Chong MM: **MicroRNA-independent roles of the RNase III enzymes Drosha and Dicer**. *Open biology* 2013, **3**(10):130144.
211. Filipowicz W, Jaskiewicz L, Kolb FA, Pillai RS: **Post-transcriptional gene silencing by siRNAs and miRNAs**. *Curr Opin Struct Biol* 2005, **15**(3):331-341.
212. Singh S, Narang AS, Mahato RI: **Subcellular fate and off-target effects of siRNA, shRNA, and miRNA**. *Pharmaceutical research* 2011, **28**(12):2996-3015.
213. Meister G, Tuschl T: **Mechanisms of gene silencing by double-stranded RNA**. *Nature* 2004, **431**(7006):343-349.
214. Malecova B, Morris KV: **Transcriptional gene silencing through epigenetic changes mediated by non-coding RNAs**. *Current opinion in molecular therapeutics* 2010, **12**(2):214-222.
215. Bayne EH, Allshire RC: **RNA-directed transcriptional gene silencing in mammals**. *Trends in genetics : TIG* 2005, **21**(7):370-373.
216. Almeida R, Allshire RC: **RNA silencing and genome regulation**. *Trends in cell biology* 2005, **15**(5):251-258.
217. Burnett JC, Rossi JJ, Tiemann K: **Current progress of siRNA/shRNA therapeutics in clinical trials**. *Biotechnology journal* 2011, **6**(9):1130-1146.
218. Kanasty R, Dorkin JR, Vegas A, Anderson D: **Delivery materials for siRNA therapeutics**. *Nature materials* 2013, **12**(11):967-977.
219. Schmidt PJ, Toudjarska I, Sendamarai AK, Racie T, Milstein S, Bettencourt BR, Hettinger J, Bumcrot D, Fleming MD: **An RNAi therapeutic targeting Tmprss6 decreases iron overload in Hfe(-/-) mice and ameliorates anemia and iron overload in murine beta-thalassemia intermedia**. *Blood* 2013, **121**(7):1200-1208.
220. Li L, Lin X, Khvorova A, Fesik SW, Shen Y: **Defining the optimal parameters for hairpin-based knockdown constructs**. *RNA (New York, NY)* 2007, **13**(10):1765-1774.
221. Xie SY, Ren ZR, Zhang JZ, Guo XB, Wang QX, Wang S, Lin D, Gong XL, Li W, Huang SZ *et al*: **Restoration of the balanced alpha/beta-globin gene expression in beta654-thalassemia mice using combined RNAi and antisense RNA approach**. *Human molecular genetics* 2007, **16**(21):2616-2625.
222. Voon HP, Wardan H, Vadolas J: **siRNA-mediated reduction of alpha-globin results in phenotypic improvements in beta-thalassemic cells**. *Haematologica* 2008, **93**(8):1238-1242.
223. Giering JC, Grimm D, Storm TA, Kay MA: **Expression of shRNA from a tissue-specific pol II promoter is an effective and safe RNAi therapeutic**. *Mol Ther* 2008, **16**(9):1630-1636.
224. Samakoglu S, Lisowski L, Budak-Alpdogan T, Usachenko Y, Acuto S, Di Marzo R, Maggio A, Zhu P, Tisdale JF, Riviere I *et al*: **A genetic strategy to treat sickle cell anemia by coregulating globin transgene expression and RNA interference**. *Nature biotechnology* 2006, **24**(1):89-94.
225. Lin SL, Chang D, Wu DY, Ying SY: **A novel RNA splicing-mediated gene silencing mechanism potential for genome evolution**. *Biochemical and biophysical research communications* 2003, **310**(3):754-760.
226. Maczuga P, Lubelski J, van Logtenstein R, Borel F, Blits B, Fakkert E, Costessi A, Butler D, van Deventer S, Petry H *et al*: **Embedding siRNA sequences targeting apolipoprotein B100 in shRNA and miRNA scaffolds results in differential processing and in vivo efficacy**. *Mol Ther* 2013, **21**(1):217-227.
227. Boudreau RL, Monteys AM, Davidson BL: **Minimizing variables among hairpin-based RNAi vectors reveals the potency of shRNAs**. *RNA (New York, NY)* 2008, **14**(9):1834-1844.
228. Fellmann C, Hoffmann T, Sridhar V, Hopfgartner B, Muhar M, Roth M, Lai DY, Barbosa IA, Kwon JS, Guan Y *et al*: **An optimized microRNA backbone for effective single-copy RNAi**. *Cell reports* 2013, **5**(6):1704-1713.
229. Xia H, Mao Q, Paulson HL, Davidson BL: **siRNA-mediated gene silencing in vitro and in vivo**. *Nature biotechnology* 2002, **20**(10):1006-1010.

230. Song J, Pang S, Lu Y, Yokoyama KK, Zheng JY, Chiu R: **Gene silencing in androgen-responsive prostate cancer cells from the tissue-specific prostate-specific antigen promoter**. *Cancer research* 2004, **64**(21):7661-7663.
231. Ren GL, Fang Y, Ma HH, Lei YF, Wang D, Xu MC, Wang PZ, Huang CX, Nie OH, Sun YT *et al*: **The short hairpin RNA driven by polymerase II suppresses both wild-type and lamivudine-resistant hepatitis B virus strains**. *Antiviral therapy* 2007, **12**(6):865-876.
232. Maczuga P, Koornneef A, Borel F, Petry H, van Deventer S, Ritsema T, Konstantinova P: **Optimization and comparison of knockdown efficacy between polymerase II expressed shRNA and artificial miRNA targeting luciferase and Apolipoprotein B100**. *BMC biotechnology* 2012, **12**:42.
233. Guda S PD, Bauer DE, Renella R, Orkin SH, Williams DA, et al: **Optimization of Lentivirus Vector RNA Polymerase II Driven microRNA Embedded shRNAs for Enhanced Processing and Efficient Knockdown of Bcl11a for Induction of Fetal Hemoglobin in Erythroid Cells**. In: *American Society of Gene & Cell Therapy, 17th Annual Meeting: 2014; Washington, U.S.A.*
234. Wilber A, Tschulena U, Hargrove PW, Kim YS, Persons DA, Barbas CF, 3rd, Nienhuis AW: **A zinc-finger transcriptional activator designed to interact with the gamma-globin gene promoters enhances fetal hemoglobin production in primary human adult erythroblasts**. *Blood* 2010, **115**(15):3033-3041.
235. Wilber A, Hargrove PW, Kim YS, Riberdy JM, Sankaran VG, Papanikolaou E, Georgomanoli M, Anagnou NP, Orkin SH, Nienhuis AW *et al*: **Therapeutic levels of fetal hemoglobin in erythroid progeny of beta-thalassemic CD34+ cells after lentiviral vector-mediated gene transfer**. *Blood* 2011, **117**(10):2817-2826.
236. Cavazzana-Calvo M, Payen E, Negre O, Wang G, Hehir K, Fusil F, Down J, Denaro M, Brady T, Westerman K *et al*: **Transfusion independence and HMGA2 activation after gene therapy of human beta-thalassaemia**. *Nature* 2010, **467**(7313):318-322.
237. Payen E, Leboulch P: **Advances in stem cell transplantation and gene therapy in the beta-hemoglobinopathies**. *Hematology / the Education Program of the American Society of Hematology American Society of Hematology Education Program* 2012, **2012**:276-283.
238. Persons DA: **Gene therapy: Targeting beta-thalassaemia**. *Nature* 2010, **467**(7313):277-278.
239. Sadelain M, Riviere I, Wang X, Boulad F, Prockop S, Giardina P, Maggio A, Galanello R, Locatelli F, Yannaki E: **Strategy for a multicenter phase I clinical trial to evaluate globin gene transfer in beta-thalassemia**. *Annals of the New York Academy of Sciences* 2010, **1202**:52-58.
240. Boulad F, Wang X, Qu J, Taylor C, Ferro L, Karponi G, Bartido S, Giardina P, Heller G, Prockop SE *et al*: **Safe mobilization of CD34+ cells in adults with beta-thalassemia and validation of effective globin gene transfer for clinical investigation**. *Blood* 2014, **123**(10):1483-1486.
241. Yannaki EaK, G.: **Current status and developments in gene therapy for thalassemia and sickle cell disease**. *Thalassemia Reports* 2014, **4**(4876).
242. Negre O, Bartholomae C, Beuzard Y, Cavazzana M, Christiansen L, Courne C, Deichmann A, Denaro M, de Dreuzy E, Finer M *et al*: **Preclinical evaluation of efficacy and safety of an improved lentiviral vector for the treatment of beta-thalassemia and sickle cell disease**. *Current gene therapy* 2015, **15**(1):64-81.
243. Laird PW, Zijderveld A, Linders K, Rudnicki MA, Jaenisch R, Berns A: **Simplified mammalian DNA isolation procedure**. *Nucleic acids research* 1991, **19**(15):4293.
244. Birnboim HC, Doly J: **A rapid alkaline extraction procedure for screening recombinant plasmid DNA**. *Nucleic acids research* 1979, **7**(6):1513-1523.
245. Wajcman H: **Analysis of hemoglobins and globin chains by high-performance liquid chromatography**. *Methods in molecular medicine* 2003, **82**:21-29.
246. Wan JH, Tian PL, Luo WH, Wu BY, Xiong F, Zhou WJ, Wei XC, Xu XM: **Rapid determination of human globin chains using reversed-phase high-performance liquid chromatography**. *Journal of chromatography B, Analytical technologies in the biomedical and life sciences* 2012, **901**:53-58.
247. Masala B, Manca L: **Detection of globin chains by reversed-phase high-performance liquid chromatography**. *Methods in enzymology* 1994, **231**:21-44.

248. **Reversed phase chromatography: Principles and methods** [http://www.google.com.cy/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&ved=0CB0QFjAA&url=http%3A%2F%2Fwolfson.huji.ac.il%2Fpurification%2FPDF%2FReversePhase%2FAmershamRPCManual.pdf&ei=8iT3U4OPIqjF0QWI34GACQ&usg=AFQjCNG3MTWiHwCl_WJryi_-7We4_eU68w]
249. **qPCR technical guide** [https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/General_Information/qpcr_technical_guide.pdf]
250. Martin P, Papayannopoulou T: **HEL cells: a new human erythroleukemia cell line with spontaneous and induced globin expression**. *Science (New York, NY)* 1982, **216**(4551):1233-1235.
251. Rutherford T, Clegg JB, Higgs DR, Jones RW, Thompson J, Weatherall DJ: **Embryonic erythroid differentiation in the human leukemic cell line K562**. *Proceedings of the National Academy of Sciences of the United States of America* 1981, **78**(1):348-352.
252. Levenson R, Housman D: **Developmental program of murine erythroleukemia cells. Effect of the inhibition of protein synthesis**. *The Journal of cell biology* 1979, **82**(3):715-725.
253. Graham FL, Smiley J, Russell WC, Nairn R: **Characteristics of a human cell line transformed by DNA from human adenovirus type 5**. *The Journal of general virology* 1977, **36**(1):59-74.
254. Bidinger B, Torres R, Rossetti RG, Brown L, Beltre R, Burstein S, Lian JB, Stein GS, Zurier RB: **Ajulemic acid, a nonpsychoactive cannabinoid acid, induces apoptosis in human T lymphocytes**. *Clinical immunology (Orlando, Fla)* 2003, **108**(2):95-102.
255. **Isolation and fractionation of mononuclear cell populations** [http://www.imba.oeaw.ac.at/uploads/media/isolation_of_mononuclear_cells_from_mouse_02.pdf]
256. **The human colony forming cells (CFC) assay using methylcellulose-based media** [http://www.rndsystems.com/literature_CFC.aspx]
257. **Hematopoietic stem cells & progenitor cells** [http://www.stemcell.com/en/Products/Cell-type/Hematopoietic-stemprogenitor-cells.aspx]
258. Sambrook J, and Russell, D. W.: **Preparation of cDNA libraries and gene identification**. In: *Molecular cloning - A laboratory manual.* vol. 2, 3rd edn. Cold Spring Harbor, N.Y., U.S.A.: Cold Spring Harbor Laboratory Press; 2001: 11.63.
259. Root DE, Hacohen N, Hahn WC, Lander ES, Sabatini DM: **Genome-scale loss-of-function screening with a lentiviral RNAi library**. *Nature methods* 2006, **3**(9):715-719.
260. Sambrook J, and Russell, D. W.: **Plasmids and their usefulness in molecular cloning**. In: *Molecular cloning - A laboratory manual.* vol. 1, 3rd edn. Cold Spring Harbor, N.Y., U.S.A.: Cold Spring Harbor Laboratory Press; 2001: 1.116.
261. Sambrook J, and Russell, D. W.: **Liquid media for E. coli**. In: *Molecular cloning - A laboratory manual.* vol. 3, 3rd edn. Cold Spring Harbor, N.Y., U.S.A.: Cold Spring Harbor Laboratory Press; 2001: A2.2.
262. Sambrook J, and Russell, D. W.: **Media containing agar or agarose**. In: *Molecular cloning - A laboratory manual.* vol. 3, 3rd edn. Cold Spring Harbor, N.Y., U.S.A.: Cold Spring Harbor Laboratory Press; 2001: A2.5.
263. Pope B, Kent HM: **High efficiency 5 min transformation of Escherichia coli**. *Nucleic acids research* 1996, **24**(3):536-537.
264. Naldini L, Blomer U, Gage FH, Trono D, Verma IM: **Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector**. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(21):11382-11388.
265. **Technical Bulletin: pAdVAntage Vector** [https://worldwide.promega.com/resources/protocols/technical-bulletins/0/padvantage-vector-protocol/]
266. Kutner RH, Zhang XY, Reiser J: **Production, concentration and titration of pseudotyped HIV-1-based lentiviral vectors**. *Nature protocols* 2009, **4**(4):495-505.
267. al Yacoub N, Romanowska M, Haritonova N, Foerster J: **Optimized production and concentration of lentiviral vectors containing large inserts**. *The journal of gene medicine* 2007, **9**(7):579-584.

268. Grigorov B, Rabilloud J, Lawrence P, Gerlier D: **Rapid titration of measles and other viruses: optimization with determination of replication cycle length**. *PloS one* 2011, **6**(9):e24135.
269. Livak KJ: **Allelic discrimination using fluorogenic probes and the 5' nuclease assay**. *Genetic analysis : biomolecular engineering* 1999, **14**(5-6):143-149.
270. Shi MM, Myrand SP, Bleavins MR, de la Iglesia FA: **High-throughput genotyping method for glutathione S-transferase T1 and M1 gene deletions using TaqMan probes**. *Research communications in molecular pathology and pharmacology* 1999, **103**(1):3-15.
271. **TaqMan Universal PCR Master Mix** [http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_042996.pdf]
272. **Creating standard curves with genomic DNA or plasmid DNA templates for use in quantitative PCR** [http://www6.appliedbiosystems.com/support/tutorials/pdf/quant_pcr.pdf]
273. Motulsky H: **Prism 5 Statistics Guide**. 2007.
274. Ghasemi A, and Zahediasl, S,. **Normality Tests for Statistical Analysis: A Guide for Non-Statisticians**. *International Journal of Endocrinology and Metabolism* 2012, **10**(2):486-489.
275. Kim HY: **Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis**. *Restorative Dentistry & Endodontics* 2013, **38**(1):52-54.
276. Dropulic B: **Lentiviral vectors: their molecular design, safety, and use in laboratory and preclinical research**. *Human gene therapy* 2011, **22**(6):649-657.
277. Escors D, Breckpot K: **Lentiviral vectors in gene therapy: their current status and future potential**. *Archivum immunologiae et therapiae experimentalis* 2010, **58**(2):107-119.
278. Goyvaerts C LT, Bricogne C, Escors D, Breckpot K. : **Targeted Lentiviral Vectors: Current Applications and Future Potential** *INTECH* 2013
279. Fang Y, Gong X, Xu M, Zeng F, Zhang J: **A self-deletion lentiviral vector to reduce the risk of replication-competent virus formation**. *The journal of gene medicine* 2013, **15**(2):102-112.
280. Pluta K, Kacprzak MM: **Use of HIV as a gene transfer vector**. *Acta biochimica Polonica* 2009, **56**(4):531-595.
281. Sakuma T, Barry MA, Ikeda Y: **Lentiviral vectors: basic to translational**. *The Biochemical journal* 2012, **443**(3):603-618.
282. Delenda C: **Lentiviral vectors: optimization of packaging, transduction and gene expression**. *The journal of gene medicine* 2004, **6 Suppl 1**:S125-138.
283. Collis P, Antoniou M, Grosveld F: **Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression**. *Embo J* 1990, **9**(1):233-240.
284. Antoniou M, Geraghty F, Hurst J, Grosveld F: **Efficient 3'-end formation of human beta-globin mRNA in vivo requires sequences within the last intron but occurs independently of the splicing reaction**. *Nucleic acids research* 1998, **26**(3):721-729.
285. Millevoi S, Geraghty F, Idowu B, Tam JL, Antoniou M, Vagner S: **A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing**. *EMBO reports* 2002, **3**(9):869-874.
286. Gelinas R, Novak U: **Retroviral vectors for the beta-globin gene that demonstrate improved titer and expression**. *Annals of the New York Academy of Sciences* 1990, **612**:427-441.
287. Miller AD, Bender MA, Harris EA, Kaleko M, Gelinas RE: **Design of retrovirus vectors for transfer and expression of the human beta-globin gene**. *J Virol* 1988, **62**(11):4337-4345.
288. Hollins C, Zorio DA, MacMorris M, Blumenthal T: **U2AF binding selects for the high conservation of the C. elegans 3' splice site**. *RNA (New York, NY)* 2005, **11**(3):248-253.
289. Nagai K, Muto Y, Pomeranz Krummel DA, Kambach C, Ignjatovic T, Walke S, Kuglstatter A: **Structure and assembly of the spliceosomal snRNPs. Novartis Medal Lecture**. *Biochemical Society transactions* 2001, **29**(Pt 2):15-26.

290. Will CL, Luhrmann R: **Spliceosome structure and function**. *Cold Spring Harb Perspect Biol* 2011, **3**(7).
291. Zhang L, Li X, Zhao R: **Structural analyses of the pre-mRNA splicing machinery**. *Protein science : a publication of the Protein Society* 2013, **22**(6):677-692.
292. Millevoi S, Decorsiere A, Loulergue C, Iacovoni J, Bernat S, Antoniou M, Vagner S: **A physical and functional link between splicing factors promotes pre-mRNA 3' end processing**. *Nucleic acids research* 2009, **37**(14):4672-4683.
293. Chen C, Zhao X, Kierzek R, Yu YT: **A flexible RNA backbone within the polypyrimidine tract is required for U2AF65 binding and pre-mRNA splicing in vivo**. *Molecular and cellular biology* 2010, **30**(17):4108-4119.
294. Singh R, Banerjee H, Green MR: **Differential recognition of the polypyrimidine-tract by the general splicing factor U2AF65 and the splicing repressor sex-lethal**. *RNA (New York, NY)* 2000, **6**(6):901-911.
295. Nelson KK, Green MR: **Mammalian U2 snRNP has a sequence-specific RNA-binding activity**. *Genes & development* 1989, **3**(10):1562-1571.
296. Madhani HD, Bordonne R, Guthrie C: **Multiple roles for U6 snRNA in the splicing pathway**. *Genes & development* 1990, **4**(12B):2264-2277.
297. Konforti BB, Konarska MM: **U4/U5/U6 snRNP recognizes the 5' splice site in the absence of U2 snRNP**. *Genes & development* 1994, **8**(16):1962-1973.
298. Vidal VP, Verdone L, Mayes AE, Beggs JD: **Characterization of U6 snRNA-protein interactions**. *RNA (New York, NY)* 1999, **5**(11):1470-1481.
299. Kambach C, Walke S, Nagai K: **Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles**. *Curr Opin Struct Biol* 1999, **9**(2):222-230.
300. Mysara M, Elhefnawi M, Garibaldi JM: **MysiRNA: improving siRNA efficacy prediction using a machine-learning model combining multi-tools and whole stacking energy (DeltaG)**. *Journal of biomedical informatics* 2012, **45**(3):528-534.
301. Li W, Cha L: **Predicting siRNA efficiency**. *Cellular and molecular life sciences : CMLS* 2007, **64**(14):1785-1792.
302. Tiscornia G, Singer O, Ikawa M, Verma IM: **A general method for gene knockdown in mice by using lentiviral vectors expressing small interfering RNA**. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(4):1844-1848.
303. Boden D, Pusch O, Silbermann R, Lee F, Tucker L, Ramratnam B: **Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins**. *Nucleic acids research* 2004, **32**(3):1154-1158.
304. Chang K, Elledge SJ, Hannon GJ: **Lessons from Nature: microRNA-based shRNA libraries**. *Nature methods* 2006, **3**(9):707-714.
305. Silva JM, Li MZ, Chang K, Ge W, Golding MC, Rickles RJ, Siolas D, Hu G, Paddison PJ, Schlabach MR *et al*: **Second-generation shRNA libraries covering the mouse and human genomes**. *Nature genetics* 2005, **37**(11):1281-1288.
306. Cullen BR: **Transcription and processing of human microRNA precursors**. *Molecular cell* 2004, **16**(6):861-865.
307. Dickins RA, Hemann MT, Zilfou JT, Simpson DR, Ibarra I, Hannon GJ, Lowe SW: **Probing tumor phenotypes using stable and regulated synthetic microRNA precursors**. *Nature genetics* 2005, **37**(11):1289-1295.
308. Stegmeier F, Hu G, Rickles RJ, Hannon GJ, Elledge SJ: **A lentiviral microRNA-based system for single-copy polymerase II-regulated RNA interference in mammalian cells**. *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(37):13212-13217.
309. McManus MT, Sharp PA: **Gene silencing in mammals by small interfering RNAs**. *Nature reviews Genetics* 2002, **3**(10):737-747.
310. Hu G, Luo J: **A primer on using pooled shRNA libraries for functional genomic screens**. *Acta biochimica et biophysica Sinica* 2012, **44**(2):103-112.
311. Ansaloni S, Lelkes N, Snyder J, Epstein C, Dubey A, Saunders AJ: **A streamlined sub-cloning procedure to transfer shRNA from a pSM2 vector to a pGIPZ lentiviral vector**. *Journal of RNAi and gene silencing : an international journal of RNA and gene targeting research* 2010, **6**(2):411-415.

312. Zhu X, Santat LA, Chang MS, Liu J, Zavzavadjian JR, Wall EA, Kivork C, Simon MI, Fraser ID: **A versatile approach to multiple gene RNA interference using microRNA-based short hairpin RNAs**. *BMC molecular biology* 2007, **8**:98.
313. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T: **Identification of novel genes coding for small expressed RNAs**. *Science (New York, NY)* 2001, **294**(5543):853-858.
314. Zeng Y, Wagner EJ, Cullen BR: **Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells**. *Molecular cell* 2002, **9**(6):1327-1333.
315. Zeng Y, Cullen BR: **Sequence requirements for micro RNA processing and function in human cells**. *RNA (New York, NY)* 2003, **9**(1):112-123.
316. Zeng Y, Yi R, Cullen BR: **Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha**. *Embo J* 2005, **24**(1):138-148.
317. Zeng Y, Cullen BR: **Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences**. *The Journal of biological chemistry* 2005, **280**(30):27595-27603.
318. Zhou H, Xia XG, Xu Z: **An RNA polymerase II construct synthesizes short-hairpin RNA with a quantitative indicator and mediates highly efficient RNAi**. *Nucleic acids research* 2005, **33**(6):e62.
319. Lebbink RJ, Lowe M, Chan T, Khine H, Wang X, McManus MT: **Polymerase II promoter strength determines efficacy of microRNA adapted shRNAs**. *PloS one* 2011, **6**(10):e26213.
320. Kunkel GR, Maser RL, Calvet JP, Pederson T: **U6 small nuclear RNA is transcribed by RNA polymerase III**. *Proceedings of the National Academy of Sciences of the United States of America* 1986, **83**(22):8575-8579.
321. Paule MR, White RJ: **Survey and summary: transcription by RNA polymerases I and III**. *Nucleic acids research* 2000, **28**(6):1283-1298.
322. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T: **Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells**. *Nature* 2001, **411**(6836):494-498.
323. Miyagishi M, Taira K: **U6 promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells**. *Nature biotechnology* 2002, **20**(5):497-500.
324. Moffat J, Grueneberg DA, Yang X, Kim SY, Kloepfer AM, Hinkle G, Piqani B, Eisenhaure TM, Luo B, Grenier JK *et al*: **A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen**. *Cell* 2006, **124**(6):1283-1298.
325. Cormack BP, Valdivia RH, Falkow S: **FACS-optimized mutants of the green fluorescent protein (GFP)**. *Gene* 1996, **173**(1 Spec No):33-38.
326. Ho HY, Cheng ML, Wang YH, Chiu DT: **Flow cytometry for assessment of the efficacy of siRNA**. *Cytometry A* 2006, **69**(10):1054-1061.
327. Yang B, Kirby S, Lewis J, Detloff PJ, Maeda N, Smithies O: **A mouse model for beta 0-thalassemia**. *Proceedings of the National Academy of Sciences of the United States of America* 1995, **92**(25):11608-11612.
328. Schaefer BC, Schaefer ML, Kappler JW, Marrack P, Kedl RM: **Observation of antigen-dependent CD8+ T-cell/ dendritic cell interactions in vivo**. *Cellular immunology* 2001, **214**(2):110-122.
329. McCulloch EA, Till JE: **The radiation sensitivity of normal mouse bone marrow cells, determined by quantitative marrow transplantation into irradiated mice**. *Radiation research* 1960, **13**:115-125.
330. Papayannopoulou TH, Lindsley D, Kurachi S, Lewison K, Hemenway T, Melis M, Anagnou NP, Najfeld V: **Adult and fetal human globin genes are expressed following chromosomal transfer into MEL cells**. *Proceedings of the National Academy of Sciences of the United States of America* 1985, **82**(3):780-784.
331. Jarvinen M: **Vimentin in human erythroleukemia (HEL) cells is modulated with differentiation inducers**. *Cell biology international reports* 1990, **14**(3):199-209.
332. Zauli G, Gibellini D, Vitale M, Secchiero P, Celeghini C, Bassini A, Pierpaoli S, Marchisio M, Guidotti L, Capitani S: **The induction of megakaryocyte differentiation**

is accompanied by selective Ser133 phosphorylation of the transcription factor CREB in both HEL cell line and primary CD34+ cells. *Blood* 1998, **92**(2):472-480.
333. Papayannopoulou T, Nakamoto B, Kurachi S, Nelson R: **Analysis of the erythroid phenotype of HEL cells: clonal variation and the effect of inducers**. *Blood* 1987, **70**(6):1764-1772.
334. Groudine M, Kohwi-Shigematsu T, Gelinas R, Stamatoyannopoulos G, Papayannopoulou T: **Human fetal to adult hemoglobin switching: changes in chromatin structure of the beta-globin gene locus**. *Proceedings of the National Academy of Sciences of the United States of America* 1983, **80**(24):7551-7555.
335. Migliaccio G, Di Pietro R, di Giacomo V, Di Baldassarre A, Migliaccio AR, Maccioni L, Galanello R, Papayannopoulou T: **In vitro mass production of human erythroid cells from the blood of normal donors and of thalassemic patients**. *Blood cells, molecules & diseases* 2002, **28**(2):169-180.
336. Kubota A, Okamura S, Shimoda K, Ikematsu W, Otsuka T, Niho Y: **Analysis of c-kit expression of human erythroleukemia cell line, HEL: clonal variation and relationship with erythroid and megakaryocytic phenotype**. *Leukemia research* 1995, **19**(4):283-290.
337. Cronin J, Zhang XY, Reiser J: **Altering the tropism of lentiviral vectors through pseudotyping**. *Current gene therapy* 2005, **5**(4):387-398.
338. Moreau-Gaudry F, Xia P, Jiang G, Perelman NP, Bauer G, Ellis J, Surinya KH, Mavilio F, Shen CK, Malik P: **High-level erythroid-specific gene expression in primary human and murine hematopoietic cells with self-inactivating lentiviral vectors**. *Blood* 2001, **98**(9):2664-2672.
339. Holmes K, Williams CM, Chapman EA, Cross MJ: **Detection of siRNA induced mRNA silencing by RT-qPCR: considerations for experimental design**. *BMC research notes* 2010, **3**:53.
340. Chen G, Kronenberger P, Teugels E, De Greve J: **Influence of RT-qPCR primer position on EGFR interference efficacy in lung cancer cells**. *Biological procedures online* 2011, **13**:1.
341. Kaufman WL, Kocman I, Agrawal V, Rahn HP, Besser D, Gossen M: **Homogeneity and persistence of transgene expression by omitting antibiotic selection in cell line isolation**. *Nucleic acids research* 2008, **36**(17):e111.
342. Enver T, Zhang JW, Anagnou NP, Stamatoyannopoulos G, Papayannopoulou T: **Developmental programs of human erythroleukemia cells: globin gene expression and methylation**. *Molecular and cellular biology* 1988, **8**(11):4917-4926.
343. Choe KS, Radparvar F, Matushansky I, Rekhtman N, Han X, Skoultchi AI: **Reversal of tumorigenicity and the block to differentiation in erythroleukemia cells by GATA-1**. *Cancer research* 2003, **63**(19):6363-6369.
344. Friend C, Scher W, Holland JG, Sato T: **Hemoglobin synthesis in murine virus-induced leukemic cells in vitro: stimulation of erythroid differentiation by dimethyl sulfoxide**. *Proceedings of the National Academy of Sciences of the United States of America* 1971, **68**(2):378-382.
345. Antoniou M: **Induction of Erythroid-Specific Expression in Murine Erythroleukemia (MEL) Cell Lines**. *Methods Mol Biol* 1991, **7**:421-434.
346. Skarpidi E, Vassilopoulos G, Stamatoyannopoulos G, Li Q: **Comparison of expression of human globin genes transferred into mouse erythroleukemia cells and in transgenic mice**. *Blood* 1998, **92**(9):3416-3421.
347. Orkin SH, Harosi FI, Leder P: **Differentiation in erythroleukemic cells and their somatic hybrids**. *Proceedings of the National Academy of Sciences of the United States of America* 1975, **72**(1):98-102.
348. Singer D, Cooper M, Maniatis GM, Marks PA, Rifkind RA: **Erythropoietic differentiation in colonies of cells transformed by Friend virus**. *Proceedings of the National Academy of Sciences of the United States of America* 1974, **71**(7):2668-2670.
349. Rovera G, Bonaiuto J: **The phenotypes of variant clones of Friend mouse erythroleukemic cells resistant to dimethyl sulfoxide**. *Cancer research* 1976, **36**(11 Pt 1):4057-4061.
350. Orlic D, Girard LJ, Anderson SM, Do BK, Seidel NE, Jordan CT, Bodine DM: **Transduction efficiency of cell lines and hematopoietic stem cells correlates with**

retrovirus receptor mRNA levels. *Stem cells (Dayton, Ohio)* 1997, **15 Suppl 1**:23-28; discussion 28-29.
351. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR *et al*: **Real-time quantification of microRNAs by stem-loop RT-PCR**. *Nucleic acids research* 2005, **33**(20):e179.
352. Varkonyi-Gasic E, Wu R, Wood M, Walton EF, Hellens RP: **Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs**. *Plant methods* 2007, **3**:12.
353. Colombo S, Nielsen HM, Foged C: **Evaluation of carrier-mediated siRNA delivery: lessons for the design of a stem-loop qPCR-based approach for quantification of intracellular full-length siRNA**. *Journal of controlled release : official journal of the Controlled Release Society* 2013, **166**(3):220-226.
354. Czimmerer Z, Hulvely J, Simandi Z, Varallyay E, Havelda Z, Szabo E, Varga A, Dezso B, Balogh M, Horvath A *et al*: **A versatile method to design stem-loop primer-based quantitative PCR assays for detecting small regulatory RNA molecules**. *PloS one* 2013, **8**(1):e55168.
355. Zhang XD: **Introduction to Genome-Scale RNAi Research**. In: *Optimal high-throughput screening: Practical experimental design and data analysis for genome-scale RNAi research.* U.S.A.: Cambridge University Press; 2011: 3-12.
356. Dighe N, Khoury M, Mattar C, Chong M, Choolani M, Chen J, Antoniou MN, Chan JK: **Long-term reproducible expression in human fetal liver hematopoietic stem cells with a UCOE-based lentiviral vector**. *PloS one* 2014, **9**(8):e104805.
357. Roelz R, Pilz IH, Mutschler M, Pahl HL: **Of mice and men: human RNA polymerase III promoter U6 is more efficient than its murine homologue for shRNA expression from a lentiviral vector in both human and murine progenitor cells**. *Experimental hematology* 2010, **38**(9):792-797.
358. Johan P: **Models of stochastic gene expression**. *Physics of Life Reviews* 2005, **2**:157-175.
359. Miller CL, Lai B: **Human and mouse hematopoietic colony-forming cell assays**. *Methods Mol Biol* 2005, **290**:71-89.
360. Magli MC, Iscove NN, Odartchenko N: **Transient nature of early haematopoietic spleen colonies**. *Nature* 1982, **295**(5849):527-529.
361. Siminovitch L, McCulloch EA, Till JE: **THE DISTRIBUTION OF COLONY-FORMING CELLS AMONG SPLEEN COLONIES**. *Journal of cellular physiology* 1963, **62**:327-336.
362. Kusser KL, Randall TD: **Simultaneous detection of EGFP and cell surface markers by fluorescence microscopy in lymphoid tissues**. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* 2003, **51**(1):5-14.
363. Tsiftsoglou AS, Vizirianakis IS, Strouboulis J: **Erythropoiesis: model systems, molecular regulators, and developmental programs**. *IUBMB life* 2009, **61**(8):800-830.
364. Weber-Benarous A, Cone RD, London IM, Mulligan RC: **Retroviral-mediated transfer and expression of human beta-globin genes in cultured murine and human erythroid cells**. *The Journal of biological chemistry* 1988, **263**(13):6142-6145.
365. Anisimov AG BI, and Volkova TO: **Exposure to sodium butyrate, dimethyl sulfoxide, and phorbol-12-myristine-13-acetate alters sensitivity of K562 cells to nonspecific lysis by human or rat leykocytes**. *Cell differentiation and proliferation* 2000, **31**(1):47-52.
366. R G: **The human erythroleukemia K562 cell culture system for identification of inducers of fetal hemoglobin**. *Minerva Biotec* 2003, **15**:123-128.
367. El-Beshlawy A, Hamdy M, El Ghamrawy M: **Fetal globin induction in beta-thalassemia**. *Hemoglobin* 2009, **33 Suppl 1**:S197-203.
368. Chan KS, Xu J, Wardan H, McColl B, Orkin S, Vadolas J: **Generation of a genomic reporter assay system for analysis of gamma- and beta-globin gene regulation**. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 2012, **26**(4):1736-1744.
369. Zhu X, Wang Y, Pi W, Liu H, Wickrema A, Tuan D: **NF-Y recruits both transcription activator and repressor to modulate tissue- and developmental stage-specific expression of human gamma-globin gene**. *PloS one* 2012, **7**(10):e47175.

370. Trakarnsanga K, Wilson MC, Lau W, Singleton BK, Parsons SF, Sakuntanaga P, Kurita R, Nakamura Y, Anstee DJ, Frayne J: **Induction of adult levels of beta-globin in human erythroid cells that intrinsically express embryonic or fetal globin by transduction with KLF1 and BCL11A-XL**. *Haematologica* 2014, **99**(11):1677-1685.
371. Chen Z, Luo HY, Steinberg MH, Chui DH: **BCL11A represses HBG transcription in K562 cells**. *Blood cells, molecules & diseases* 2009, **42**(2):144-149.
372. Papayannopoulou T, Nakamoto B, Yokochi T, Chait A, Kannagi R: **Human erythroleukemia cell line (HEL) undergoes a drastic macrophage-like shift with TPA**. *Blood* 1983, **62**(4):832-845.
373. Poncz M, Surrey S, LaRocco P, Weiss MJ, Rappaport EF, Conway TM, Schwartz E: **Cloning and characterization of platelet factor 4 cDNA derived from a human erythroleukemic cell line**. *Blood* 1987, **69**(1):219-223.
374. Gui CY, Jiang C, Xie HY, Qian RL: **The apoptosis of HEL cells induced by hydroxyurea**. *Cell research* 1997, **7**(1):91-97.
375. Agyekum D WM, and Meiler SE.: **Resveratrol inhibits GATA-1 and BCL11A protein expression and increases fetal hemoglobin in K562 erythroleukemia cells**. In: *SCDAA 39th Annual Convention; Memphis, Tennessee*. 2011.
376. Agyekum D CZ, Xiao H, Meiler SE, et a: **Resveratrol Regulates Fetal Hemoglobin Production by Inhibiting KLF1 and BCL11A in KU812 Cells** In: *54th ASH Annual Meeting and Exposition; Atlanta, GA*. 2012.
377. Pourfarzad F, von Lindern M, Azarkeivan A, Hou J, Kia SK, Esteghamat F, van Ijcken W, Philipsen S, Najmabadi H, Grosveld F: **Hydroxyurea responsiveness in beta-thalassemic patients is determined by the stress response adaptation of erythroid progenitors and their differentiation propensity**. *Haematologica* 2013, **98**(5):696-704.
378. Endo T, Ishibashi Y, Okana H, Fukumaki Y: **Significance of pH on differentiation of human erythroid cell lines**. *Leukemia research* 1994, **18**(1):49-54.
379. Talbot D, Collis P, Antoniou M, Vidal M, Grosveld F, Greaves DR: **A dominant control region from the human beta-globin locus conferring integration site-independent gene expression**. *Nature* 1989, **338**(6213):352-355.
380. MacNeill EC, Hanenberg H, Pollok KE, van der Loo JC, Bierhuizen MF, Wagemaker G, Williams DA: **Simultaneous infection with retroviruses pseudotyped with different envelope proteins bypasses viral receptor interference associated with colocalization of gp70 and target cells on fibronectin CH-296**. *J Virol* 1999, **73**(5):3960-3967.
381. Wotherspoon S, Dolnikov A, Symonds G, Nordon R: **Susceptibility of cell populations to transduction by retroviral vectors**. *J Virol* 2004, **78**(10):5097-5102.
382. Park JI, Choi HS, Jeong JS, Han JY, Kim IH: **Involvement of p38 kinase in hydroxyurea-induced differentiation of K562 cells**. *Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research* 2001, **12**(9):481-486.
383. Samid D, Yeh A, Prasanna P: **Induction of erythroid differentiation and fetal hemoglobin production in human leukemic cells treated with phenylacetate**. *Blood* 1992, **80**(6):1576-1581.
384. Trakarnsanga K, Wilson MC, Lau W, Singleton BK, Parsons SF, Sakuntanaga P, Kurita R, Nakamura Y, Anstee DJ, Frayne J: **Induction of adult levels of beta-globin in human erythroid cells that intrinsically express embryonic or fetal globin by transduction with KLF1 and BCL11A-XL**. *Haematologica* 2014.
385. Ho AD, Punzel M: **Hematopoietic stem cells: can old cells learn new tricks?** *Journal of leukocyte biology* 2003, **73**(5):547-555.
386. Narsinh KH, Plews, J. and Wu, J.C.: **Comparison of Human Induced Pluripotent and Embryonic Stem Cells: Fraternal or Identical Twins?** *Molecular Therapy* 2011, **19**(4).
387. Gardner RL: **Stem cells: potency, plasticity and public perception**. *Journal of anatomy* 2002, **200**(Pt 3):277-282.
388. Thoma SJ, Lamping CP, Ziegler BL: **Phenotype analysis of hematopoietic CD34+ cell populations derived from human umbilical cord blood using flow cytometry and cDNA-polymerase chain reaction**. *Blood* 1994, **83**(8):2103-2114.

389. Hogan CJ, Shpall EJ, Keller G: **Differential long-term and multilineage engraftment potential from subfractions of human CD34+ cord blood cells transplanted into NOD/SCID mice**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(1):413-418.
390. Genovese P, Schiroli G, Escobar G, Di Tomaso T, Firrito C, Calabria A, Moi D, Mazzieri R, Bonini C, Holmes MC *et al*: **Targeted genome editing in human repopulating haematopoietic stem cells**. *Nature* 2014, **510**(7504):235-240.
391. Pulsipher MA, Chitphakdithai P, Logan BR, Shaw BE, Wingard JR, Lazarus HM, Waller EK, Seftel M, Stroncek DF, Lopez AM *et al*: **Acute toxicities of unrelated bone marrow versus peripheral blood stem cell donation: results of a prospective trial from the National Marrow Donor Program**. *Blood* 2013, **121**(1):197-206.
392. Nagler A, Labopin M, Shimoni A, Mufti GJ, Cornelissen JJ, Blaise D, Janssen JJ, Milpied N, Vindelov L, Petersen E *et al*: **Mobilized peripheral blood stem cells compared with bone marrow from HLA-identical siblings for reduced-intensity conditioning transplantation in acute myeloid leukemia in complete remission: a retrospective analysis from the Acute Leukemia Working Party of EBMT**. *European journal of haematology* 2012, **89**(3):206-213.
393. Lannert H, Able T, Becker S, Sommer M, Braun M, Stadtherr P, Ho AD: **Optimizing BM harvesting from normal adult donors**. *Bone marrow transplantation* 2008, **42**(7):443-447.
394. Yannaki E, Karponi G, Zervou F, Constantinou V, Bouinta A, Tachynopoulou V, Kotta K, Jonlin E, Papayannopoulou T, Anagnostopoulos A *et al*: **Hematopoietic stem cell mobilization for gene therapy: superior mobilization by the combination of granulocyte-colony stimulating factor plus plerixafor in patients with beta-thalassemia major**. *Human gene therapy* 2013, **24**(10):852-860.
395. Frittoli MC, Biral E, Cappelli B, Zambelli M, Roncarolo MG, Ferrari G, Ciceri F, Marktel S: **Bone marrow as a source of hematopoietic stem cells for human gene therapy of beta-thalassemia**. *Human gene therapy* 2011, **22**(4):507-513.
396. Sirinoglu Demiriz I, Tekgunduz E, Altuntas F: **What is the most appropriate source for hematopoietic stem cell transplantation? Peripheral stem cell/bone marrow/cord blood**. *Bone marrow research* 2012, **2012**:834040.
397. Jin H, Kim HS, Kim S, Kim HO: **Erythropoietic potential of CD34+ hematopoietic stem cells from human cord blood and G-CSF-mobilized peripheral blood**. *BioMed research international* 2014, **2014**:435215.
398. Smith AR, Wagner JE: **Alternative haematopoietic stem cell sources for transplantation: place of umbilical cord blood**. *British journal of haematology* 2009, **147**(2):246-261.
399. Hoggatt J, Pelus LM: **Mobilization of hematopoietic stem cells from the bone marrow niche to the blood compartment**. *Stem cell research & therapy* 2011, **2**(2):13.
400. Korbling M: **Peripheral Blood Stem Cells: A Novel Source for Allogeneic Transplantation**. *The oncologist* 1997, **2**(2):104-113.
401. Alwasaidi T, and Bredeson, C.: **Peripheral blood stem cells or bone marrow as the graft source for allogeneic hematopoietic cell transplantation?** *Journa of Taibah University Medical Sciences* 2014, **9**(2):91 - 99.
402. Zhang M, Huang B: **The multi-differentiation potential of peripheral blood mononuclear cells**. *Stem cell research & therapy* 2012, **3**(6):48.
403. Boehm D, Murphy WG, Al-Rubeai M: **The potential of human peripheral blood derived CD34+ cells for ex vivo red blood cell production**. *Journal of biotechnology* 2009, **144**(2):127-134.
404. Masiello F, Tirelli V, Sanchez M, van den Akker E, Gabriella G, Marconi M, Villa MA, Rebulla P, Hashmi G, Whitsett C *et al*: **Mononuclear cells from a rare blood donor, after freezing under good manufacturing practice conditions, generate red blood cells that recapitulate the rare blood phenotype**. *Transfusion* 2014, **54**(4):1059-1070.
405. Pan Q, Ye L, Deng Z, Li L, Liu H: **Effects of red blood cell lysing solutions on the detection of peripheral basophils of healthy normals and SLE patients by flow cytometry**. *Journal of immunoassay & immunochemistry* 2014, **35**(4):368-377.

406. Bossuyt X, Marti GE, Fleisher TA: **Comparative analysis of whole blood lysis methods for flow cytometry**. *Cytometry* 1997, **30**(3):124-133.
407. Quillen K, Berkman EM: **Methods of isolation and cryopreservation of stem cells from cord blood**. *Journal of hematotherapy* 1996, **5**(2):153-155.
408. Mallone R, Mannering SI, Brooks-Worrell BM, Durinovic-Bello I, Cilio CM, Wong FS, Schloot NC: **Isolation and preservation of peripheral blood mononuclear cells for analysis of islet antigen-reactive T cell responses: position statement of the T-Cell Workshop Committee of the Immunology of Diabetes Society**. *Clinical and experimental immunology* 2011, **163**(1):33-49.
409. Klein AB, Witonsky SG, Ahmed SA, Holladay SD, Gogal RM, Jr., Link L, Reilly CM: **Impact of different cell isolation techniques on lymphocyte viability and function**. *Journal of immunoassay & immunochemistry* 2006, **27**(1):61-76.
410. Einwallner E, Subasic A, Strasser A, Augustin D, Thalhammer R, Steiner I, Schwarzinger I: **Lysis matters: red cell lysis with FACS Lyse affects the flow cytometric enumeration of circulating leukemic blasts**. *Journal of immunological methods* 2013, **390**(1-2):127-132.
411. Tiirikainen MI: **Evaluation of red blood cell lysing solutions for the detection of intracellular antigens by flow cytometry**. *Cytometry* 1995, **20**(4):341-348.
412. Yeo C, Saunders N, Locca D, Flett A, Preston M, Brookman P, Davy B, Mathur A, Agrawal S: **Ficoll-Paque versus Lymphoprep: a comparative study of two density gradient media for therapeutic bone marrow mononuclear cell preparations**. *Regenerative medicine* 2009, **4**(5):689-696.
413. Chang Y, Hsieh PH, Chao CC: **The efficiency of Percoll and Ficoll density gradient media in the isolation of marrow derived human mesenchymal stem cells with osteogenic potential**. *Chang Gung medical journal* 2009, **32**(3):264-275.
414. Posel C, Moller K, Frohlich W, Schulz I, Boltze J, Wagner DC: **Density gradient centrifugation compromises bone marrow mononuclear cell yield**. *PloS one* 2012, **7**(12):e50293.
415. De Paoli P, Reitano M, Battistin S, Castiglia C, Santini G: **Enumeration of human lymphocyte subsets by monoclonal antibodies and flow cytometry: a comparative study using whole blood or mononuclear cells separated by density gradient centrifugation**. *Journal of immunological methods* 1984, **72**(2):349-353.
416. Tamul KR, Schmitz JL, Kane K, Folds JD: **Comparison of the effects of Ficoll-Hypaque separation and whole blood lysis on results of immunophenotypic analysis of blood and bone marrow samples from patients with hematologic malignancies**. *Clinical and diagnostic laboratory immunology* 1995, **2**(3):337-342.
417. Tamul KR, O'Gorman MR, Donovan M, Schmitz JL, Folds JD: **Comparison of a lysed whole blood method to purified cell preparations for lymphocyte immunophenotyping: differences between healthy controls and HIV-positive specimens**. *Journal of immunological methods* 1994, **167**(1-2):237-243.
418. Romeu MA, Mestre M, Gonzalez L, Valls A, Verdaguer J, Corominas M, Bas J, Massip E, Buendia E: **Lymphocyte immunophenotyping by flow cytometry in normal adults. Comparison of fresh whole blood lysis technique, Ficoll-Paque separation and cryopreservation**. *Journal of immunological methods* 1992, **154**(1):7-10.
419. Ashmore LM, Shopp GM, Edwards BS: **Lymphocyte subset analysis by flow cytometry. Comparison of three different staining techniques and effects of blood storage**. *Journal of immunological methods* 1989, **118**(2):209-215.
420. Renzi P, Ginns LC: **Analysis of T cell subsets in normal adults. Comparison of whole blood lysis technique to Ficoll-Hypaque separation by flow cytometry**. *Journal of immunological methods* 1987, **98**(1):53-56.
421. Toth TE, Smith B, Pyle H: **Simultaneous separation and purification of mononuclear and polymorphonuclear cells from the peripheral blood of cats**. *Journal of virological methods* 1992, **36**(2):185-195.
422. Strasser A, Kalmar E, Niedermuller H: **A simple method for the simultaneous separation of peripheral blood mononuclear and polymorphonuclear cells in the dog**. *Veterinary immunology and immunopathology* 1998, **62**(1):29-35.
423. Sastry L, Johnson T, Hobson MJ, Smucker B, Cornetta K: **Titering lentiviral vectors: comparison of DNA, RNA and marker expression methods**. *Gene therapy* 2002, **9**(17):1155-1162.

424. Charde MS IP, Welankiwar AS, Kumar J, and Chakole RD: **Review: The procurement, storage and quality assurance of frozen blood and tissue biospecimens**. *International Journal of Pharmacological Research* 2014, **4**(2).
425. Stevens VL, Patel AV, Feigelson HS, Rodriguez C, Thun MJ, Calle EE: **Cryopreservation of whole blood samples collected in the field for a large epidemiologic study**. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2007, **16**(10):2160-2163.
426. Germann A, Schulz JC, Kemp-Kamke B, Zimmermann H, von Briesen H: **Standardized Serum-Free Cryomedia Maintain Peripheral Blood Mononuclear Cell Viability, Recovery, and Antigen-Specific T-Cell Response Compared to Fetal Calf Serum-Based Medium**. *Biopreservation and biobanking* 2011, **9**(3):229-236.
427. van den Akker E, Satchwell TJ, Pellegrin S, Daniels G, Toye AM: **The majority of the in vitro erythroid expansion potential resides in CD34(-) cells, outweighing the contribution of CD34(+) cells and significantly increasing the erythroblast yield from peripheral blood samples**. *Haematologica* 2010, **95**(9):1594-1598.
428. Selkirk SM: **Gene therapy in clinical medicine**. *Postgraduate medical journal* 2004, **80**(948):560-570.
429. Heike T, Nakahata T: **Ex vivo expansion of hematopoietic stem cells by cytokines**. *Biochimica et biophysica acta* 2002, **1592**(3):313-321.
430. Glimm H, Oh IH, Eaves CJ: **Human hematopoietic stem cells stimulated to proliferate in vitro lose engraftment potential during their S/G(2)/M transit and do not reenter G(0)**. *Blood* 2000, **96**(13):4185-4193.
431. Lewis PF, Emerman M: **Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus**. *J Virol* 1994, **68**(1):510-516.
432. Bukrinsky MI, Haggerty S, Dempsey MP, Sharova N, Adzhubel A, Spitz L, Lewis P, Goldfarb D, Emerman M, Stevenson M: **A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells**. *Nature* 1993, **365**(6447):666-669.
433. Wang CX, Sather BD, Wang X, Adair J, Khan I, Singh S, Lang S, Adams A, Curinga G, Kiem HP *et al*: **Rapamycin relieves lentiviral vector transduction resistance in human and mouse hematopoietic stem cells**. *Blood* 2014, **124**(6):913-923.
434. Santoni de Sio FR, Cascio P, Zingale A, Gasparini M, Naldini L: **Proteasome activity restricts lentiviral gene transfer into hematopoietic stem cells and is down-regulated by cytokines that enhance transduction**. *Blood* 2006, **107**(11):4257-4265.
435. Santoni de Sio F, Naldini L: **Short-term culture of human CD34+ cells for lentiviral gene transfer**. *Methods Mol Biol* 2009, **506**:59-70.
436. Kennedy DR, McLellan K, Moore PF, Henthorn PS, Felsburg PJ: **Effect of ex vivo culture of CD34+ bone marrow cells on immune reconstitution of XSCID dogs following allogeneic bone marrow transplantation**. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation* 2009, **15**(6):662-670.
437. **Guidance for Industry: Guidance for Human Somatic Cell Therapy and Gene Therapy** [http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/Guidances/CellularandGeneTherapy/ucm072987.htm#viii]
438. E. F: **The use of cell culture procedures for studying fetal hemoglobin stimulating drugs**. *Minerva Biotecnologica* 2003, **15**(2):129-136.
439. Ronzoni L, Bonara P, Rusconi D, Frugoni C, Libani I, Cappellini MD: **Erythroid differentiation and maturation from peripheral CD34+ cells in liquid culture: cellular and molecular characterization**. *Blood cells, molecules & diseases* 2008, **40**(2):148-155.
440. Malik P, Fisher TC, Barsky LL, Zeng L, Izadi P, Hiti AL, Weinberg KI, Coates TD, Meiselman HJ, Kohn DB: **An in vitro model of human red blood cell production from hematopoietic progenitor cells**. *Blood* 1998, **91**(8):2664-2671.
441. Zhang CC, Lodish HF: **Cytokines regulating hematopoietic stem cell function**. *Current opinion in hematology* 2008, **15**(4):307-311.

442. Panzenbock B, Bartunek P, Mapara MY, Zenke M: **Growth and differentiation of human stem cell factor/erythropoietin-dependent erythroid progenitor cells in vitro**. *Blood* 1998, **92**(10):3658-3668.
443. Giarratana MC, Kobari L, Lapillonne H, Chalmers D, Kiger L, Cynober T, Marden MC, Wajcman H, Douay L: **Ex vivo generation of fully mature human red blood cells from hematopoietic stem cells**. *Nature biotechnology* 2005, **23**(1):69-74.
444. Brugger W, Mocklin W, Heimfeld S, Berenson RJ, Mertelsmann R, Kanz L: **Ex vivo expansion of enriched peripheral blood CD34+ progenitor cells by stem cell factor, interleukin-1 beta (IL-1 beta), IL-6, IL-3, interferon-gamma, and erythropoietin**. *Blood* 1993, **81**(10):2579-2584.
445. Dorn I, Lazar-Karsten P, Boie S, Ribbat J, Hartwig D, Driller B, Kirchner H, Schlenke P: **In vitro proliferation and differentiation of human CD34+ cells from peripheral blood into mature red blood cells with two different cell culture systems**. *Transfusion* 2008, **48**(6):1122-1132.
446. Carlile GW, Smith DH, Wiedmann M: **Caspase-3 has a nonapoptotic function in erythroid maturation**. *Blood* 2004, **103**(11):4310-4316.
447. Miharada K, Hiroyama T, Sudo K, Nagasawa T, Nakamura Y: **Efficient enucleation of erythroblasts differentiated in vitro from hematopoietic stem and progenitor cells**. *Nature biotechnology* 2006, **24**(10):1255-1256.
448. Fujimi A, Matsunaga T, Kobune M, Kawano Y, Nagaya T, Tanaka I, Iyama S, Hayashi T, Sato T, Miyanishi K *et al*: **Ex vivo large-scale generation of human red blood cells from cord blood CD34+ cells by co-culturing with macrophages**. *International journal of hematology* 2008, **87**(4):339-350.
449. Migliaccio G, Migliaccio AR, Druzin ML, Giardina PJ, Zsebo KM, Adamson JW: **Long-term generation of colony-forming cells in liquid culture of CD34+ cord blood cells in the presence of recombinant human stem cell factor**. *Blood* 1992, **79**(10):2620-2627.
450. Goodman JW, Hall EA, Miller KL, Shinpock SG: **Interleukin 3 promotes erythroid burst formation in "serum-free" cultures without detectable erythropoietin**. *Proceedings of the National Academy of Sciences of the United States of America* 1985, **82**(10):3291-3295.
451. Millington M, Arndt A, Boyd M, Applegate T, Shen S: **Towards a clinically relevant lentiviral transduction protocol for primary human CD34 hematopoietic stem/progenitor cells**. *PloS one* 2009, **4**(7):e6461.
452. **Frequencies of cell types in human peripheral blood** [http://www.stemcell.com/~/media/Files/wallchart_CellTypes_WEB.pdf]
453. Fibach E, Kollia P, Schechter AN, Noguchi CT, Rodgers GP: **Hemin-induced acceleration of hemoglobin production in immature cultured erythroid cells: preferential enhancement of fetal hemoglobin**. *Blood* 1995, **85**(10):2967-2974.
454. Fibach E, Manor D, Oppenheim A, Rachmilewitz EA: **Proliferation and maturation of human erythroid progenitors in liquid culture**. *Blood* 1989, **73**(1):100-103.
455. Breda L, Kleinert DA, Casu C, Casula L, Cartegni L, Fibach E, Mancini I, Giardina PJ, Gambari R, Rivella S: **A preclinical approach for gene therapy of beta-thalassemia**. *Annals of the New York Academy of Sciences* 2010, **1202**:134-140.
456. Li J, Hale J, Bhagia P, Xue F, Chen L, Jaffray J, Yan H, Lane J, Gallagher PG, Mohandas N *et al*: **Isolation and transcriptome analyses of human erythroid progenitors: BFU-E and CFU-E**. *Blood* 2014, **124**(24):3636-3645.
457. Fibach E, Burke LP, Schechter AN, Noguchi CT, Rodgers GP: **Hydroxyurea increases fetal hemoglobin in cultured erythroid cells derived from normal individuals and patients with sickle cell anemia or beta-thalassemia**. *Blood* 1993, **81**(6):1630-1635.
458. Kustikova OS, Wahlers A, Kuhlcke K, Stahle B, Zander AR, Baum C, Fehse B: **Dose finding with retroviral vectors: correlation of retroviral vector copy numbers in single cells with gene transfer efficiency in a cell population**. *Blood* 2003, **102**(12):3934-3937.
459. Charrier S, Ferrand M, Zerbato M, Precigout G, Viornery A, Bucher-Laurent S, Benkhelifa-Ziyyat S, Merten OW, Perea J, Galy A: **Quantification of lentiviral vector copy numbers in individual hematopoietic colony-forming cells shows vector dose-dependent effects on the frequency and level of transduction**. *Gene therapy* 2011, **18**(5):479-487.

460. Romero Z, Urbinati F, Geiger S, Cooper AR, Wherley J, Kaufman ML, Hollis RP, de Assin RR, Senadheera S, Sahagian A *et al*: **beta-globin gene transfer to human bone marrow for sickle cell disease**. *The Journal of clinical investigation* 2013.
461. Wong ML, Medrano JF: **Real-time PCR for mRNA quantitation**. *BioTechniques* 2005, **39**(1):75-85.
462. Ghielmini M, Pfister U, Zucca E, van den Bosch S, Tamasy P, Marangoni G, Bertoli G, Derivaz JL, Cavalli F: **Distribution of mobilized progenitor cells in the buffy coat of the haemonetics MCS3p cell separator: a study to optimize the collection of progenitors by leukapheresis**. *Journal of hematotherapy* 1998, **7**(3):251-256.
463. Meyer TP, Zehnter I, Hofmann B, Zaisserer J, Burkhart J, Rapp S, Weinauer F, Schmitz J, Illert WE: **Filter Buffy Coats (FBC): a source of peripheral blood leukocytes recovered from leukocyte depletion filters**. *Journal of immunological methods* 2005, **307**(1-2):150-166.
464. Dow LE, Premsrirut PK, Zuber J, Fellmann C, McJunkin K, Miething C, Park Y, Dickins RA, Hannon GJ, Lowe SW: **A pipeline for the generation of shRNA transgenic mice**. *Nature protocols* 2012, **7**(2):374-393.
465. Peixeiro I, Silva AL, Romao L: **Control of human beta-globin mRNA stability and its impact on beta-thalassemia phenotype**. *Haematologica* 2011, **96**(6):905-913.
466. Brendel C GS, Renella R, Du P, Bauer DE, Canver MC, Kamran SC, Thornton J, de Boer H, Milsom MD, Orkin SH, Gregory R, Williams DA: **Optimization of Bcl11a Knockdown By miRNA Scaffold Embedded Shrnas Leading to Enhanced Induction of Fetal Hemoglobin in Erythroid Cells for the Treatment of Beta-Hemoglobinopathies**. In: *56th ASH Annual Meeting and Exposition: December 6-9; San Francisco, CA*. 2014.
467. Khaled WT, Choon Lee S, Stingl J, Chen X, Raza Ali H, Rueda OM, Hadi F, Wang J, Yu Y, Chin SF *et al*: **BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells**. *Nature communications* 2015, **6**:5987.
468. Auyeung VC, Ulitsky I, McGeary SE, Bartel DP: **Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing**. *Cell* 2013, **152**(4):844-858.
469. Myburgh R, Cherpin O, Schlaepfer E, Rehrauer H, Speck RF, Krause KH, Salmon P: **Optimization of Critical Hairpin Features Allows miRNA-based Gene Knockdown Upon Single-copy Transduction**. *Molecular therapy Nucleic acids* 2014, **3**:e207.
470. Bartel DP: **MicroRNAs: target recognition and regulatory functions**. *Cell* 2009, **136**(2):215-233.
471. Lulli V, Romania P, Morsilli O, Cianciulli P, Gabbianelli M, Testa U, Giuliani A, Marziali G: **MicroRNA-486-3p regulates gamma-globin expression in human erythroid cells by directly modulating BCL11A**. *PloS one* 2013, **8**(4):e60436.
472. Jiang BY, Zhang XC, Su J, Meng W, Yang XN, Yang JJ, Zhou Q, Chen ZY, Chen ZH, Xie Z *et al*: **BCL11A overexpression predicts survival and relapse in non-small cell lung cancer and is modulated by microRNA-30a and gene amplification**. *Molecular cancer* 2013, **12**:61.
473. Sander JD, Joung JK: **CRISPR-Cas systems for editing, regulating and targeting genomes**. *Nature biotechnology* 2014, **32**(4):347-355.